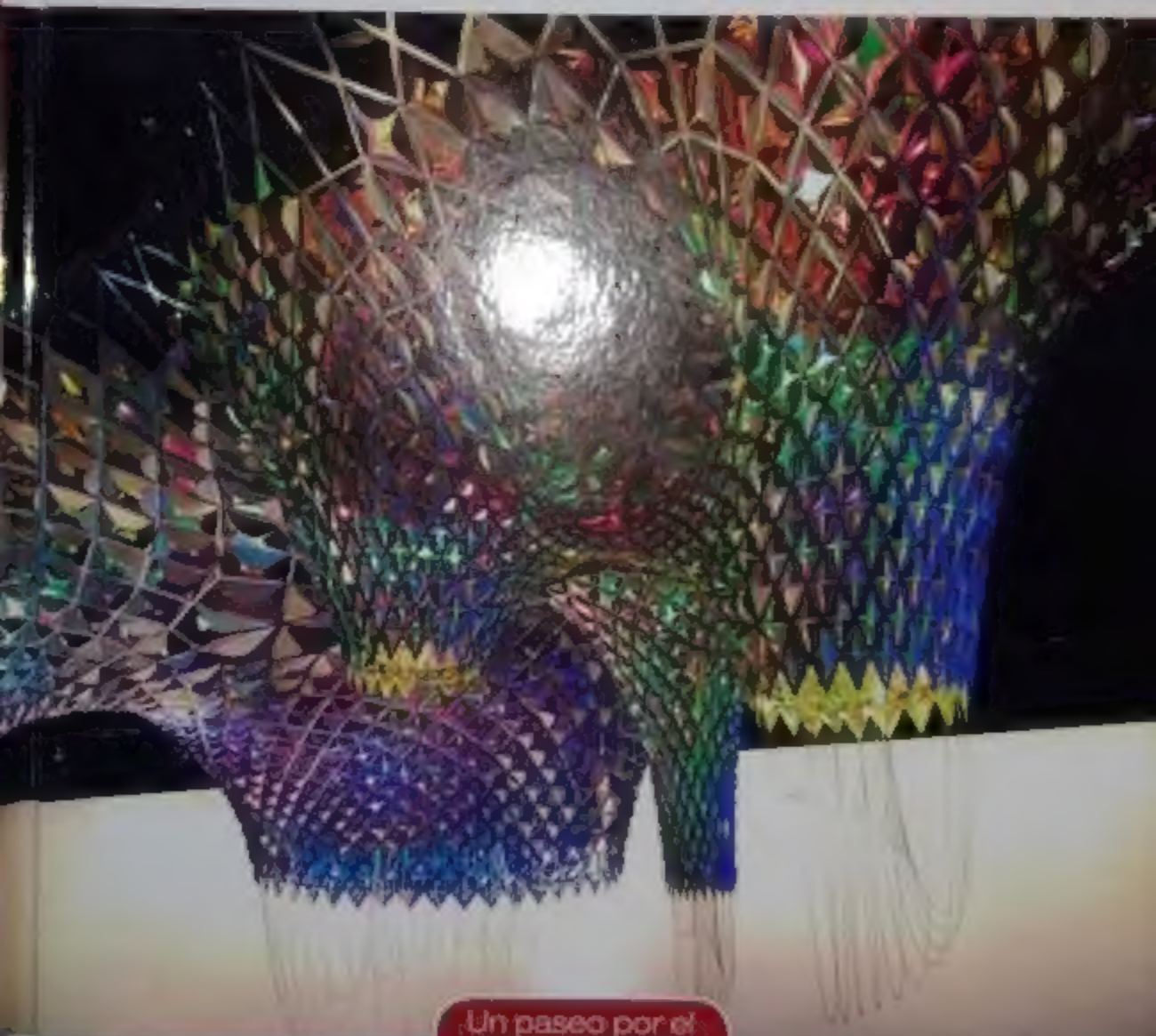


La inteligencia artificial

De los circuitos al conocimiento



Un paseo por el
COSMOS

EXLIBRIS Scan Digit



The Doctor y La Comunidad

Redigitalización: The Doctor

<http://thedoctorwho1967.blogspot.com.ar/>

<http://el1900.blogspot.com.ar/>

<http://librosrevistasinteresesanexo.blogspot.com.ar/>

<https://labibliotecadeldrmourea.blogspot.com/>

La inteligencia artificial

De los circuitos al conocimiento

RBA

Imagen de cubierta: Escultura flotante diseñada por Watson (sistema informático de inteligencia artificial desarrollado por IBM) e inspirada en la arquitectura gaudiniana. La escultura, presentada en el Mobile World Congress de 2017, cambiaba de forma y color en función de las emociones expresadas en las redes sociales sobre una serie de temas escogidos para el evento.

Dirección científica de la colección: Manuel Lozano Leyva

© Ignasi Belda Reig por el texto
 © RBA Contenidos Editoriales y Audiovisuales, S.A.U.
 © 2017, RBA Coleccionables, S.A.

Realización: EDITEC

Diseño cubierta: Llorenç Martí

Diseño interior: tactilestudio

Infografías: Joan Pejoan

Fotografías: Arabella Im/Wikimedia Commons: 29b; Archivo RBA: 19, 23, 27, 32, 135b; Ignasi Belda: portada; Criteria/Wikimedia Commons: 30; Dangleis/Wikimedia Commons: 67ad; DARPA: 130; Mary Evans Picture Library Ltd/Age Fotostock: 21; Rick Friedman/Getty Images: 89; Google: 113; Google/Getty Images: 123b; Herman Hiddema/Wikimedia Commons: 123a; Stan Honda/Getty Images: 93a, 93b; Konstantin Lanzet/Wikimedia Commons: 25; Frank Mayfield/Wikimedia Commons: 67b; Night Wolf Dezines/Vexels: 76; NVIDIA Corporation: 119; Oleksiy/123RF: 61b; Bill Pierce/The LIFE Images Collection/Getty Images: 61a; Radomil/Wikimedia Commons: 67ai; Science Museum London/Science and Society Picture Library: 29a; Kurt Seebauer/Wikimedia Commons: 48; Christiaan Dirk Tonnis: 51; Michael L. Umbrecht y Carl R. Friend/Wikimedia Commons: 87; Universidad de Tübinga: 135a, 135c.

Reservados todos los derechos. Ninguna parte de esta publicación puede ser reproducida, almacenada o transmitida por ningún medio sin permiso del editor.

ISBN: 978-84-473-8877-6

Depósito legal: B-8279-2017

Impreso y encuadernado en Rodesa, Villatuerta (Navarra)

Impreso en España - Printed in Spain

SUMARIO

INTRODUCCIÓN	7
CAPÍTULO 1	La prehistoria de la inteligencia artificial 13
CAPÍTULO 2	Con Turing se hizo la luz 35
CAPÍTULO 3	El avance de la inteligencia artificial y las guerras 55
CAPÍTULO 4	La inteligencia artificial en el presente 83
CAPÍTULO 5	Big data en el día a día 103
CAPÍTULO 6	Algunas disquisiciones desde la ética y la filosofía 125
LECTURAS RECOMENDADAS	145
ÍNDICE	147

INTRODUCCIÓN

¿Qué es la inteligencia artificial? Esta es la pregunta que, a lo largo del último siglo, matemáticos, ingenieros y filósofos han intentado contestar no sin un amplio debate. Desde ciertos sectores a menudo la cuestión se ha planteado de forma sesgada. Por ejemplo, Hollywood ha presentado reiteradamente conceptos muy alejados de la realidad presente, pasada y futura. Solo hace falta visionar la película *2001: Una odisea del espacio*, dirigida por Stanley Kubrick y basada en la novela de Arthur C. Clarke de 1968, para comprobar que, más de una década después de ese 2001, la inteligencia artificial aún está lejos del imaginario popular que el filme contribuyó a crear.

La complejidad epistemológica de la pregunta se asemeja a la presentada por otras cuestiones similares como ¿qué es la inteligencia? o, incluso, ¿qué es la vida? De hecho, filósofos y biólogos aún no han consensuado una respuesta, que además de difícil está sometida al inevitable sesgo antropocéntrico. Cuando un humano tiene que definir qué son la vida o la inteligencia, es inevitable que su contestación se haga desde una óptica humana, la cual obvia otros tipos de vidas posibles u otros tipos de inteligencia.

A título de ejemplo, en psicología existen corrientes de pensamiento que defienden que la inteligencia humana está compuesta por una serie de inteligencias múltiples que se superponen, entre ellas la inteligencia lingüística, la lógico-matemática, la espacial, la musical, la corporal, la intrapersonal, etc. El balance de todas es lo que le confiere inteligencia al ser humano. Sin embargo, no está claro cuál es el balance óptimo entre ellas y ni siquiera cómo medirlas. ¿Podemos decir que la estrella del baloncesto Michael Jordan es más inteligente que el más célebre de los físicos, Albert Einstein? No hay duda de que uno posee una inteligencia corporal que supera con creces la del otro y viceversa en cuanto a inteligencia lógico-matemática, pero ¿quién es el más inteligente?

Si el debate de la inteligencia humana aún no está resuelto, podemos suponer que el de la inteligencia artificial todavía lo está menos. Pero para complicarlo todo un poco más, ¿qué pasa con las inteligencias colectivas de determinados tipos de colonias animales? Existen muchas cuestiones aún por solventar sobre las inteligencias colectivas que presentan las comunidades de hormigas o las bandadas de aves que migran cada año de forma perfectamente coordinada y con un gran sentido de la orientación.

Una de las muchas características cognitivas que distingue un computador de una mente humana es la gran capacidad de cálculo del primero en comparación con la segunda. Por ejemplo, un simple móvil actual tiene una potencia de cálculo de unos 25 GFLOPS, lo que vienen a ser unos 25 000 millones de operaciones (sumas, restas, multiplicaciones, etc.) en coma flotante, es decir, con decimales, por segundo. ¿Cuántas operaciones de este tipo puede hacer un humano por segundo? ¿Podemos decir que un móvil es por ello más inteligente que un humano? Obviamente no, pero tampoco podemos dejar que la perspectiva antropocéntrica nos determine la definición de la inteligencia. Y lo mismo pasa si hablamos de memoria. ¿Cuántos miles de billones de datos puede almacenar un computador doméstico actual? ¿Puede la mente humana almacenar una milésima parte de esos datos?

Otra característica clave en la definición de inteligencia es la autoconsciencia, aquella que nos permite ser sabedores de nues-

tra propia existencia. ¿Es la autoconsciencia el hecho que define a un ser inteligente? ¿Tienen las hormigas, poseedoras de una inteligencia colectiva, autoconsciencia? ¿Cómo podemos saber si un organismo o artefacto es autoconsciente? De hecho, un sistema con inteligencia artificial podría estar programado para responder «sí» ante la pregunta ¿eres tú autoconsciente?, con lo cual es imposible determinar si realmente un sistema es autoconsciente o está simplemente mintiendo. La autoconsciencia es, además, la capacidad de reconocerse a uno mismo y, por ende, de reconocer cada uno de los miembros del grupo como una individualidad. Con la autoconsciencia surgen sentimientos como la trascendencia, la compasión o el altruismo. Con la autoconsciencia las relaciones sociales se estrechan y se hacen más complejas. Mediante la adquisición de la consciencia del yo, el *Homo sapiens*, por primera vez, adquiere la habilidad de reconocer que es un individuo independiente y diferente de los otros. Aparece la capacidad de representarse mentalmente a uno mismo y todo esto permite que se asignen diferentes papeles a los miembros del grupo, haciendo posible el mantenimiento de los individuos más débiles. Finalmente, como consecuencia de todo ello, aparecen las creencias del más allá para soportar y explicar el concepto de la muerte y, por tanto, el desarrollo de las ideas religiosas y mágicas. En definitiva, la autoconsciencia es lo que nos hace humanos, pero ¿es un requisito *sine qua non* para ser inteligente? Desde un punto de vista antropocéntrico la respuesta debería ser afirmativa, pero ¿estaríamos en lo cierto?

Por otra parte, hay quien opina que la inteligencia se define mediante la creatividad. Un humano es un ser creativo pero no parece que un ordenador pueda serlo. Sin embargo, esta realidad está cambiando, y en la última década hemos visto programas de ordenador que componen piezas musicales o incluso pintan cuadros imitando el estilo de Van Gogh. ¿Son estos algoritmos inteligentes?

Quizá los tiros no vayan por la vía de la autoconsciencia, ni tampoco tengan que ver con la creatividad, la potencia de cálculo o la capacidad de almacenar datos, sino con la aptitud de percibir sensaciones, es decir, por el desarrollo de los sentidos

y la capacidad de interrelacionar datos externos con una base de conocimiento y poder expresar unas conclusiones, de forma colectiva o individual. La psicología moderna defiende que el ser humano dispone de nueve sentidos, en contraposición a los cinco sentidos clásicos; estos son la vista, el olfato, el tacto, el gusto, el oído, la percepción térmica, el sentido del dolor fisiológico o nocicepción, el equilibrio y la propiocepción o sentido kinestésico.

La propiocepción es un sentido interesante, pues se relaciona con la autoconsciencia. Es la percepción del conocimiento del cuerpo y de la situación de sus diferentes partes, desde la punta del dedo del pie hasta la uña de cualquier dedo de las manos. Es más, si cogemos un destornillador para atornillar un mueble, nuestra propiocepción se extenderá hasta la punta de la herramienta. Lo mismo pasa al conducir un coche e intentar aparcarlo: la propiocepción se amplía a todos los límites del coche, de tal forma que podemos aparcarlo sin rozar con otros elementos externos. Un computador ¿puede tener propiocepción?, ¿puede tener autoconsciencia de todas sus teclas? La intuición nos dice que no, pero la experiencia nos demuestra que hoy en día existen muchos coches comerciales que se aparcan solos. ¿Acaso los ordenadores de a bordo están dotados de propiocepción?

Como la propia semántica indica, los sentidos están ahí porque tenemos sensores. Por ejemplo, el sentido del equilibrio se percibe desde un sensor orgánico llamado cóclea o caracol. Este, a través de un complejo sistema de canales, puede percibir la posición relativa del individuo en relación al centro de gravedad de la Tierra. Sin embargo, ¿cuántos sensores tiene un teléfono móvil convencional? Docenas, entre los cuales se hallan los sensores giroscópicos, que dotan de equilibrio al teléfono móvil y permiten rotar la orientación de la pantalla en función de la posición del aparato, sea apaisada o vertical.

Llegados a este punto queda claro que no existe aún una respuesta sencilla ante la pregunta ¿qué es la inteligencia? Y si dicha pregunta ya es de respuesta polémica y esquiva, en el presente libro daremos una nueva vuelta de tuerca e intentaremos responder a la pregunta ¿qué es la inteligencia artificial? Como veremos a lo largo de estas páginas, el célebre matemático y filósofo Alan

Turing será una figura clave en el intento de dar una respuesta a la cuestión. Sin embargo, tenemos que reconocer que, a pesar de los grandes avances conceptuales introducidos por Turing, su visión era completamente antropocéntrica y, por tanto, en sus definiciones de inteligencia no entrarían las complejas colonias de insectos que usan su inteligencia colectiva para la planificación del futuro u otro tipo de inteligencias artificiales que la ciencia ficción no se cansa de relatarnos.

Hoy en día la inteligencia artificial nos rodea y está presente en todas las actividades diarias que realizamos. Uno de los mayores hitos tecnológicos resueltos por los ingenieros en inteligencia artificial ha sido conseguir que esta sea completamente transparente al usuario y que haya penetrado en nuestras vidas sin darnos cuenta.

En este libro veremos los fascinantes cambios que han ido modelando esta disciplina científica, evolucionando en tan solo unos años desde los inventos más rudimentarios hasta los sistemas capaces de controlar nuestras vidas. Explicaremos los inicios de dicha disciplina, sus figuras clave, sus hitos más notables y la manera en que se ha ido introduciendo en nuestra cotidianeidad hasta el punto de resultar inherente al modo de vida del hombre occidental actual. Quizá entonces, asimilado todo este conocimiento, tal vez podamos vislumbrar una respuesta a nuestras anteriores preguntas y conocernos un poco más, tanto a nosotros mismos como a nuestras compañeras inseparables: las máquinas.

La prehistoria de la inteligencia artificial

En la época anterior al célebre Alan Turing, la historia de la inteligencia artificial se confunde con la de las ciencias de la computación. Ese periodo está repleto de anécdotas y de sorprendentes artefactos que fueron irrumpiendo en una sociedad cada vez más fascinada por los automatismos.

Aunque parezca inverosímil, las raíces de la inteligencia artificial y, por ende, de las ciencias de la computación, son tan profundas que se remontan a la antigua Grecia. Cerca del año 300 a.C., el sabio Aristóteles (384 a.C.-322 a.C.), prolífico autor de unos doscientos tratados sobre temas de lo más diverso —desde la lógica a la filosofía, pasando por la política, la biología, las matemáticas o la astronomía—, fundamentó un sistema lógico denominado *silogismo*, que se considera el germen de la lógica moderna. La lógica es el concepto fundamental mediante el cual se sustenta la inteligencia artificial; de ahí la importancia de este primitivo precedente: un sistema inteligente destinado a desarrollar razonamientos que, lejos de seguir procesos o reglas aleatorias, utiliza algún tipo de lógica para ir deduciendo o infiriendo conclusiones.

En concreto, el silogismo es una forma de razonamiento deductivo constituido por dos proposiciones que actúan como premisas y por una tercera, la conclusión, que es una inferencia deductiva de las dos primeras. Este sistema es un método lógico que relaciona de forma ordenada dos términos diferentes, un sujeto y un predicado, que tienen en común un término medio. Las conclusiones se obtienen mediante unas reglas denomina-

das «juicios». El sujeto es la «premisa menor», y el predicado, la «premisa mayor». Por tanto, el juicio permite conjugar una combinación de ambas para determinar una conclusión. De esta forma, la lógica silogística ofrece leyes, concretamente diecinueve, que permiten garantizar la veracidad de las conclusiones.

Algunas de ellas pueden recordarse mediante unas sencillas reglas mnemotécnicas ideadas ya en época moderna: Bárbara (tres «a»), Celarent (una «e», una «a» y otra «e»), Darii (una «a» y dos «ies») o Ferio (una «e», una «i» y una «o»). Las palabras indican a través de sus vocales las cláusulas que engloba: una «a» si el predicado es universal afirmativo, una «e» si es universal negativo, una «i» si es particular afirmativo y una «o» si es particular negativo. Pongamos como ejemplo un silogismo del tipo Celarent:

Predicado: No todos los niños son traviesos.

Sujeto: Todos los estudiantes de esta escuela son niños.

Conclusión: No todos los estudiantes de esta escuela son traviesos.

El Celarent está compuesto de una cláusula universal negativa —no todos los niños son traviesos—, otra universal afirmativa —todos los estudiantes de esta escuela son niños— y una conclusión que es universal negativa —no todos los estudiantes de esta escuela son traviesos—.

TRAS ARISTÓTELES, CTESIBIO DE ALEJANDRÍA

Avanzando en la línea del tiempo, el siguiente pensador que no podemos olvidar en la prehistoria de la inteligencia artificial es Ctesibio de Alejandría (285 a.C.-222 a.C.), un gran científico e ingeniero griego de la Antigüedad del que, desgraciadamente, no se ha conservado ninguno de sus tratados, aunque por suerte, Ateneo de Náucratis, gramático griego de principios del siglo III a.C., compiló íntegramente sus trabajos. A Ctesibio se le atribuye el gran mérito de ser el precursor de la neumática, ya que fue el primero en tratar y estudiar la ciencia de la compre-

sión de gases, su elasticidad y sus implicaciones en el desarrollo de bombas (para bombear líquidos) y cañones. Pero, además, fue un gran experto en hidráulica: desarrolló diferentes bombas hidráulicas de gran potencia, el órgano acuático e, incluso, el sifón. Sin embargo, para el tema que nos ocupa, lo más destacable de este sabio es que, cerca de 250 a.C., construyó la primera máquina autocontrolada, concretamente un regulador del flujo del agua que actuaba siguiendo un método programado. Podríamos decir que esta máquina es el primer algoritmo de la historia de la humanidad, pues por vez primera el hombre había logrado construir un artefacto que se comportaba siguiendo unas instrucciones introducidas de antemano en el sistema.

Se trata de la clepsidra o reloj de agua, un mecanismo que, mediante la regulación de un flujo de agua, permite medir el tiempo transcurrido (véase la imagen de la pág. 19). Parece que la clepsidra desarrollada por Ctesibio tenía mayor precisión que cualquiera de los relojes anteriores y posteriores, y que no fue superada hasta que en el siglo XVII el astrónomo y físico holandés Christiaan Huygens (1629-1695) inventó el péndulo. Por tanto, podemos afirmar que Aristóteles y él fueron los dos precursores de las ciencias de la computación. Mientras que el primero inventó la lógica primitiva, al otro se le atribuye la algorítmica.

AVANCES MEDIEVALES EN PRO DE LA INTELIGENCIA ARTIFICIAL

Ciertamente la Edad Media fue una época oscura en términos de producción científica, pero existen algunos pocos destellos que, de forma inequívoca, contribuyeron al desarrollo de la inteligencia artificial tal y como la conocemos hoy en día.

Nacen las bases de datos: Isidoro de Sevilla

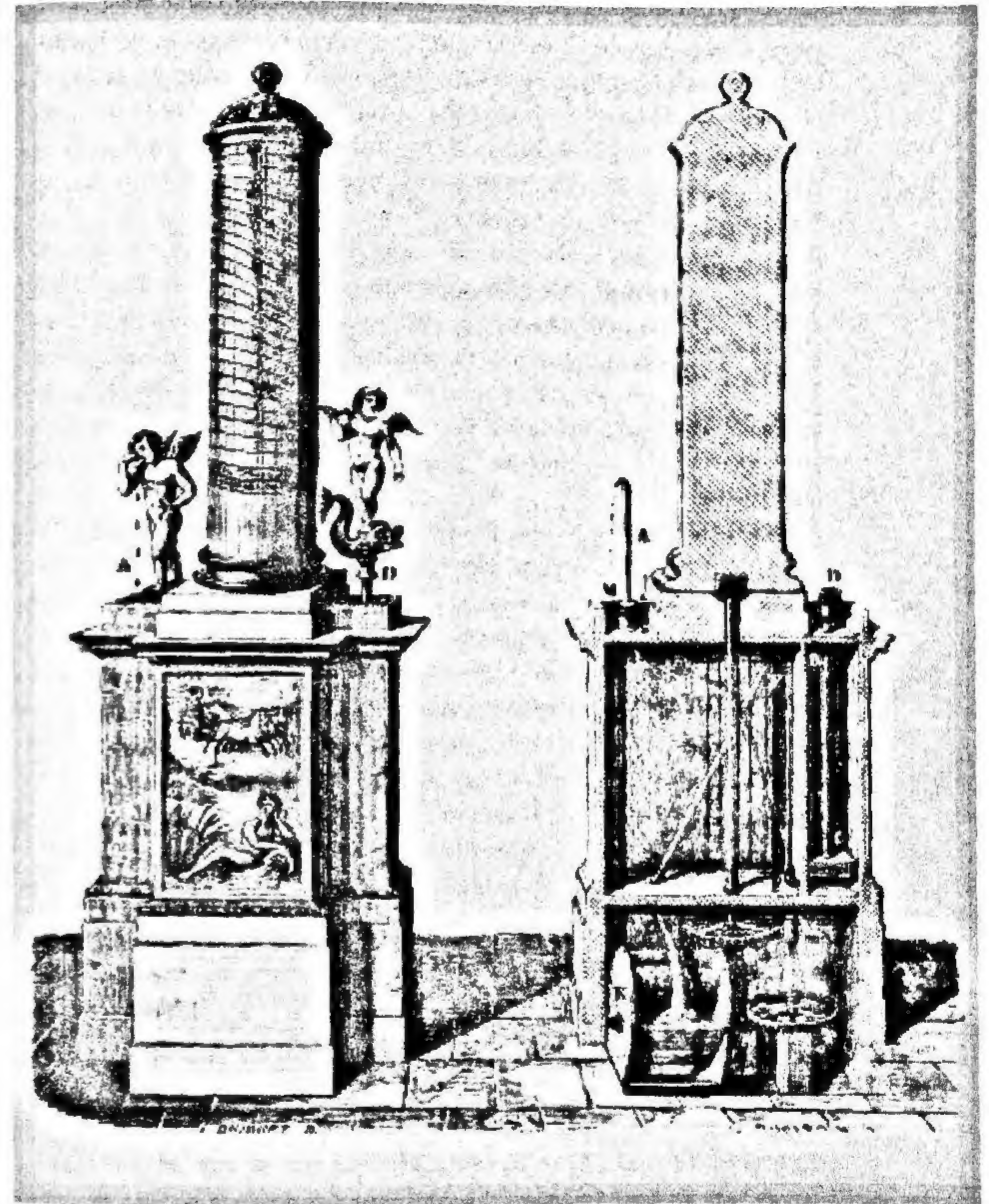
El primero de ellos vino de la mano del eclesiástico erudito Isidoro de Sevilla (556-636), el primero de los grandes compiladores medievales, canonizado en 1598 (véase la imagen de la pág. 21).

Fue él quien, sin saberlo, estableció los principios modernos de las actuales bases de datos mientras acometía esas compilaciones. Precisamente por esos inventarios de conocimiento, en 2003 el santo fue propuesto como el patrón de internet. De hecho, en algunos lugares se celebra el 4 de abril el Día de Internet en memoria de la fecha de su defunción. Porque en realidad ¿qué es una base de datos sino un sistema de conocimiento que aglutina de forma ordenada y coherente los hechos conocidos sobre un determinado entorno? Un ejemplo podría ser la creación de la base de datos de las películas que contiene nuestra colección particular. Los campos que puede tener dicha base de datos podrían ser: Título, Director, Actor Principal, Nacionalidad, Duración, Año de Producción. A partir de aquí, la base de datos es una tabla con una columna para cada campo y tantas filas como películas tengamos almacenadas.

Obviamente, mucho ha llovido desde que san Isidoro planteó estas compilaciones, sembrando la génesis de esas bases de datos arcaicas que, con el tiempo, se fueron organizando alrededor de estructuras más complejas. Primero fueron las bases de datos jerárquicas, más tarde las relacionales y, en la actualidad, los sistemas de gestión de bases de datos *NoSQL*, útiles cuando se trabaja en entornos de datos masivos o *big data*. Los sistemas de conocimiento que ayudan a estructurar las bases de datos son imprescindibles para que un algoritmo inteligente pueda aprender, razonar y, en definitiva, tomar decisiones en entornos complejos.

El nacimiento del algoritmo

Tras este santo, hoy patrón de las redes de comunicación del siglo XIX, la siguiente aportación relevante en la prehistoria de la inteligencia artificial y, por extensión, de las ciencias de la computación, fue Al-Juarismi (ca. 780-850). A este matemático y pensador persa medieval se le atribuye la invención del algoritmo, que no es más que una sucesión de instrucciones concretas, no ambiguas, ordenadas y finitas, que se ejecutan ante una entrada dada para generar una salida concreta. Pondremos un



Clepsidra desarrollada por Ctesibio según el arquitecto Claude Perrault. También llamado reloj de agua, este mecanismo mide el paso del tiempo en función de lo que tarda el agua en pasar de un recipiente a otro.

ejemplo que es habitual al iniciar el primer día de clase de los estudios universitarios de ciencias de la computación: pedirles a los alumnos que desarrollen el algoritmo del huevo frito. Hagámoslo nosotros:

1. Abrir la nevera.
2. Seleccionar un huevo.
3. Encender la cocina.
4. Colocar una sartén sobre el fuego.
5. Verter aceite dentro de la sartén.
6. Esperar a que el aceite esté a la temperatura adecuada.
7. Echar el contenido del huevo sobre el aceite.
8. Esparcir sal sobre el huevo.
9. Esperar a que el huevo esté cocinado.
10. Apagar el fuego.
11. Disponer el contenido de la sartén sobre un plato.

Aunque quizá no lo parezca, escribir un algoritmo sobre un tema tan trivial como freír un huevo no es nada obvio para un principiante. De hecho, el error más habitual que comete un programador novel cuando se le pide escribir este algoritmo es el de realizar tareas mientras el aceite se calienta —por ejemplo, ir a buscar el huevo a la nevera—, algo imposible en la programación imperativa y secuencial con la que los computadores trabajan.

Sin embargo, no hay que desanimarse. Iniciarse en el «arte» de la programación de computadores no es misión imposible. Después de un poco de práctica, cualquiera puede desarrollar correctamente algoritmos de casi cualquier cosa, por ejemplo, sobre cómo ordenar una serie de números aleatorios, de menor a mayor, como en el siguiente ejemplo:

3, 7, 34, 1, 3, 98, 5 → algoritmo de ordenación → 1, 3, 3, 5, 7, 34, 98.

Al-Juarismi se interesó también por la astronomía y la geografía, pero sobre todo por el álgebra; de hecho, se le considera el padre de esta rama de las matemáticas. Su obra principal es *Hi-*



Desde 2003 san Isidoro es considerado el patrón de internet e incluso existe una plegaria al respecto: «Te suplicamos que, por la intercesión de san Isidoro, obispo y doctor, durante nuestros viajes a través de internet dirijamos nuestras manos y ojos solo para lo que es agradable a Ti y tratar con caridad y paciencia todas esas almas que nos encontramos».

sāb al-ġabr wa'l muqābala, traducida al español como *Compendio de cálculo por compleción y comparación*, escrita entre 813 y 833 d.C. Fue aquí donde estableció los cimientos del álgebra y, de hecho, a esta obra le debemos los términos «álgebra» o «algoritmo». De esta obra solo se conserva una copia de 1361, escrita en árabe, custodiada en la Universidad de Oxford, aunque más adelante se publicaron traducciones al inglés y otras lenguas, por lo que afortunadamente no se ha perdido. Pero Al-Juarismi no solo sentó las bases del álgebra, disciplina que estudia la combinación de elementos de estructuras abstractas según determinadas reglas y que es la herramienta que nos permite resolver ecuaciones. También propuso diversos algoritmos para resolver problemas algebraicos.

Un gran sabio inspirado e iluminado

Unos quinientos años después del sabio musulmán, nacía el que es considerado como el gran sabio medieval de Occidente, el mallorquín Ramon Llull o Raimundo Lulio (1232-1316), como también se le conoce en castellano. Llull fue un filósofo, poeta y científico mallorquín al que le debemos multitud de avances y conceptos en el plano intelectual. Su enorme obra, que abarca desde tratados de filosofía o teología hasta novelas, pasando por la poesía o la divulgación científica, ha llegado hasta nuestros días intacta, escrita en catalán, árabe y latín. Llull, conocido en su época con los apodos de *Doctor Inspiratus* y *Doctor Illuminatus*, escribió 243 libros, y su influencia en sabios, filósofos y matemáticos posteriores ha sido caudal, desde Leibniz, Pascal o Boole en el plano abstracto-matemático, hasta un gran número de autores en el ámbito más filosófico y teológico.

Pero lo que es relevante respecto al tema de la inteligencia artificial es el ingenio mecánico que construyó, al que bautizó como *Ars Magna* (Gran Arte). Era una especie de máquina pensante que, mediante diales, palancas y manivelas, era capaz de realizar demostraciones lógicas en base a sujetos y predicados de teorías teológicas que la máquina combinaba para compro-

EL ARS MAGNA DE RAMON LLULL

El impacto de Ramon Llull en la inteligencia artificial pasa por la construcción de su *Ars Magna* o máquina lógica, un artefacto de naturaleza mecánica construido para probar por sí mismo la naturaleza verdadera o falsa de un postulado. Para ello Llull inventó un sistema lógico evolucionado del aristotélico. La lógica de Llull permite combinar de forma correcta «principios absolutos» con «principios correlativos», a través de los «principios relativos», en relación a los «sujetos», de acuerdo con unas «reglas interrogativas». De hecho, la innovación de la máquina lógica de Llull era que, mecánicamente, implementaba estas reglas de forma combinatoria; reglas que en la lógica aristotélica equivaldrían a las ya mencionadas Bárbara, Celarent, Darii, Ferio, etcétera.



De lo divino a lo racional

En el *Ars Magna*, los sujetos y predicados lógicos estaban organizados en una serie de figuras geométricas perfectas, tales como círculos, cuadrados, triángulos, etc. Por tanto, el aparato, operado mediante unos diales y palancas, era capaz de situar las proposiciones y tesis a lo largo de unos viales hasta la posición de verdadero o de falso, según correspondiera. En otras palabras, el *Ars Magna* de Ramon Llull implementaba un sistema lógico operado de forma combinatoria y automática, lo cual era una gran innovación en pleno siglo XIII. Su funcionamiento se basaba en la combinación de diversos aspectos divinos para llegar a conclusiones racionales. Dichos aspectos divinos, codificados mediante diferentes letras de la B a la K, se podían combinar mediante unas figuras de escaleras y círculos.

Razonamiento mecánico

Cabe destacar que los esfuerzos de Ramon Llull por construir un sistema lógico-filosófico perfecto e irrefutable tenían un trasfondo teológico. Su objetivo último era romper las fronteras intelectuales entre judíos, cristianos y musulmanes y establecer una terminología aceptable para todos. Su voluntad pasaba por construir un sistema de razonamiento automático —en este caso pasando por la mecánica— mediante el cual se pudieran obtener conclusiones verdaderas en los campos de la filosofía y la religión sin la barrera de las lenguas y vocablos concretos. Al contrario que otras escuelas filosóficas contemporáneas, como la escolástica, Llull descartó los textos sagrados u otros dogmas de fe para la obtención de sus conclusiones. Por tanto, su sistema se basaba única y exclusivamente en una argumentación racional. El hecho de haber podido implementar esta argumentación en un artificio mecánico imposibilitaba introducir valoraciones subjetivas u otro tipo de sesgos humanos o incluso irracionales.

bar, según Llull, si eran ciertas o falsas. En la actualidad, es considerado el patrono de la ingeniería informática.

SIGLO XIX: EMPIEZA UNA REVOLUCIÓN CIENTÍFICA

Ya lo vimos con Aristóteles: la lógica matemática es una parte de las matemáticas y la filosofía que se encarga de estudiar la forma del razonamiento. Es decir, es una disciplina que, por medio de reglas y técnicas, determina si un argumento es válido o no. La lógica es el fundamento sobre el cual se sustentan las ciencias de la computación y la inteligencia artificial, y se utiliza como instrumento para validar o derivar nuevo conocimiento. Fueron dos británicos decimonónicos, George Boole y Augustus De Morgan, los que, con sus leyes lógicas, dieron una nueva vuelta de tuerca a las bases de la lógica aristotélica y, mediante una nueva notación más abstracta, la llevaron hasta el siguiente nivel, el de la lógica de primer orden, en contraposición a la lógica clásica de Aristóteles. Este estadio implica solo expresiones finitas y fórmulas bien definidas, y no deja lugar a dominios infinitos ni incertidumbres. A pesar de ser tan estricta, la lógica de primer orden ofrece una gran potencia simbólica que, hasta los días actuales, permite la resolución de complejos problemas de razonamiento. Mientras que el primero de estos dos personajes, el matemático y filósofo británico George Boole (1815-1864), desarrolló el álgebra de Boole o booleana, base de la aritmética computacional moderna en la que está fundamentada toda la electrónica digital, el segundo, el matemático británico de origen indio Augustus De Morgan (1806-1871), ideó un sistema de reglas que permiten expresar, manipular y simplificar problemas lógicos que admiten dos estados, verdadero o falso, por procedimientos matemáticos.

Veamos algunos ejemplos simples de esta lógica de primer orden. Para entenderlos hay que introducir primero los símbolos más usuales de la lógica, que son la negación (\neg), la conjunción (\wedge), la disyunción (\vee) y la implicación (\rightarrow). De forma adicional, Boole introdujo un sistema algebraico que facilitaba la manipu-

ÁLGBRA BOOLEANA

La denominada álgebra booleana se basa en tres operaciones matemáticas básicas: la negación o complemento, la disyunción u «o» y la conjunción o «y». La negación, representada con el símbolo \neg , invierte el estado de una variable. Por ejemplo si $A = \text{«Aristóteles es un hombre»}$, entonces $\neg A = \text{«Aristóteles no es un hombre»}$. La disyunción, representada con el símbolo \vee , es un operador binario, es decir, necesita dos parámetros para obtener un resultado. Este es verdadero siempre que alguno de los dos parámetros sea cierto. Por ejemplo, ¿es verdadero que lo que usted está haciendo ahora es leer «o» conducir? La respuesta es que sí, es verdadero, dado que lo que ahora está haciendo es leer este

libro. Si usted estuviera ahora conduciendo y no leyendo, la respuesta a la pregunta también sería cierta, e incluso, sería igualmente cierta si estuviera cometiendo la grave imprudencia de leer este libro y conducir al mismo tiempo. Finalmente, el tercer operador es la intersección, representada con el símbolo \wedge y que también es un operador binario. Si reformulamos la pregunta anterior y preguntamos, ¿es verdadero que lo que usted está haciendo ahora es leer «y» conducir?, únicamente podríamos decir que sí en el caso de que estuviera cometiendo tamaño disparate. A partir de estos tres operadores se pueden construir otros más sofisticados, como la o-exclusiva ($\bar{\wedge}$), que solo sería cierta, en el ejemplo anterior, si estamos leyendo o conduciendo, pero no si hacemos las dos cosas a la vez. El operador $\bar{\wedge}$ no es un operador básico del álgebra booleana, dado que se puede reescribir mediante los otros tres operadores básicos: $A \bar{\wedge} B = (A \wedge \neg B) \vee (\neg A \wedge B)$.

Avances espectaculares

En electrónica digital el significado de verdadero y falso se simboliza con un 1 y un 0 respectivamente. El 1 representa pasar la corriente eléctrica y el 0 no dejarla pasar. Las operaciones lógicas se realizan mediante la combinación de transistores y la alta integración de estos dispositivos electrónicos ha permitido que la informática haya avanzado tanto (en la imagen un microprocesador de 1999) en los últimos cuarenta años. El microprocesador del ordenador de su casa, que es el «cerebro central» del aparato, contiene centenares de millones de transistores en pocos centímetros cuadrados.

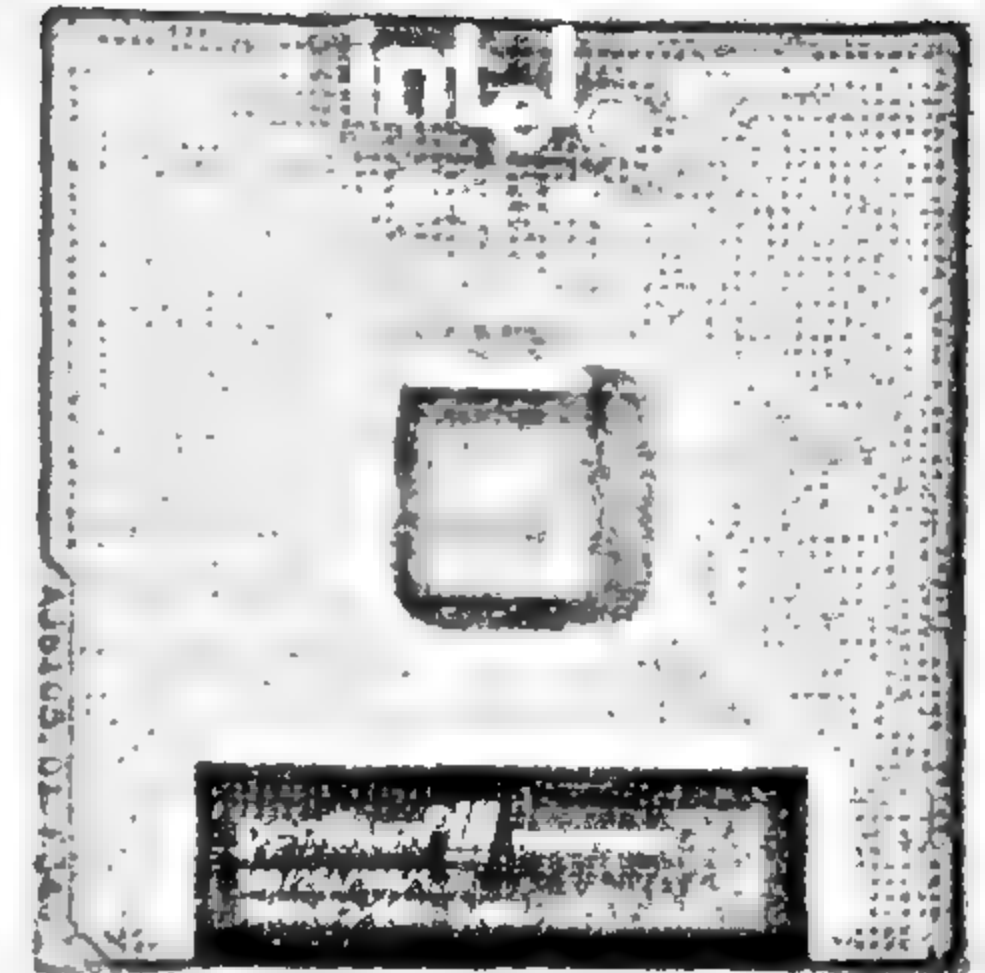


Foto del popular microprocesador de Intel, Pentium III.

lación de variables. Por ejemplo, si decimos que me llamo Ignacio, podemos decir que el *axioma* «me llamo Ignacio» es A. De esta forma, $\neg A$ sería el equivalente a «no me llamo Ignacio». Con todo esto, veamos el primer ejemplo: $A \wedge \neg A \rightarrow \text{FALSO}$. Es decir, «me llamo Ignacio» y «no me llamo Ignacio», necesariamente tiene que ser falso. Sin embargo, el segundo ejemplo, $A \vee \neg A \rightarrow \text{VERDADERO}$, es decir, «me llamo Ignacio» o «no me llamo Ignacio» necesariamente tiene que ser verdadero, ya que algún nombre hay que tener, ya sea Ignacio u otro.

Pongamos un tercer ejemplo: si $A \rightarrow B$ y $B \rightarrow C$, entonces $A \rightarrow C$. Traslademos el álgebra a palabras. Pongamos por caso que el axioma A es «yo», el axioma B es «humano» y el axioma C es «mamífero». Podríamos leer la secuencia algebraica « $A \rightarrow B$ y $B \rightarrow C$, entonces $A \rightarrow C$ »: «como yo soy humano y los humanos son mamíferos, por tanto, yo soy un mamífero». O un cuarto: si $A \rightarrow B$ y A es VERDADERO, entonces B también es VERDADERO. Veámoslo expresado en palabras. Al axioma A le asignamos el valor «enero», mientras que B es «frío». Por tanto, $A \rightarrow B$ nos dice que en enero hace frío, al menos en el hemisferio norte. Ahora decimos que A es VERDADERO, o sea que hoy estamos en enero. Por tanto, B también es VERDADERO, es decir, que hace frío.

A esta lógica booleana Augustus De Morgan le añadió un potente sistema deductivo —o reglas de derivación— para ir combinando las reglas lógicas e ir evolucionando en el razonamiento lógico. Por ejemplo, se puede demostrar que $\neg(A \wedge B) = \neg A \vee \neg B$ y, de hecho, esta equivalencia se conoce como *regla de De Morgan*. Aunque, para ser justos, De Morgan no fue el único que trabajó en ampliar la lógica booleana, sino que también cabe destacar otros matemáticos ingleses y estadounidenses, como William Stanley Jevons, Charles Sanders Peirce y William Ernest Johnson.

Con todos estos avances, hoy en día tenemos una lógica de primer orden que, como decíamos antes, es aún la piedra angular de la informática moderna, la inteligencia artificial y el razonamiento automático. Por ejemplo, en el año 1956, el economista Herbert Simon junto con el experto en ciencias de la computación Allen Newell, ambos estadounidenses, desarro-

LAS LEYES DE DE MORGAN

Ampliamente reconocido por sus aportaciones en la lógica matemática y, en particular, por establecer las bases de la inducción matemática (el método de manipulación algebraica que dota de potencia a la lógica de primer orden), el matemático británico Augustus De Morgan hizo otra gran aportación: las leyes que llevan su nombre y que permiten transformar disyunciones en conjunciones y viceversa. Una de las reglas indica que $\neg(A \wedge B) = \neg A \vee \neg B$ y la otra que $\neg(A \vee B) = \neg A \wedge \neg B$. Vamos a demostrar la primera de ellas mediante el método conocido como la tabla de la verdad (las abreviaciones V y F corresponden a VERDADERO y FALSO, respectivamente).

A	B	$A \wedge B$	$\neg(A \wedge B)$	$\neg A$	$\neg B$	$\neg A \vee \neg B$
V	V	V	F	F	F	F
V	F	F	V	F	V	V
F	V	F	V	V	F	V
F	F	F	V	V	V	V

Como se puede observar, las columnas sombreadas, la de $\neg(A \wedge B)$ y la de $\neg A \vee \neg B$, tienen la misma secuencia de V y F, es decir, si (primera fila) A es VERDADERO y B es VERDADERO, el resultado de $\neg(A \wedge B)$ como de $\neg A \vee \neg B$ es siempre FALSO. Lo mismo pasa si A es VERDADERO y B FALSO, si A es FALSO y B VERDADERO, o si A es FALSO y B también. Como solo hay dos variables, A y B, únicamente existen cuatro combinaciones posibles. Para todas ellas el resultado de las dos expresiones algebraicas es idéntico. Por tanto, queda demostrada la veracidad de esta regla propuesta por De Morgan.

Esta imagen de alrededor de 1860 muestra al matemático, filósofo, profesor y abogado Augustus de Morgan cuando tenía unos cincuenta y cinco años de edad.

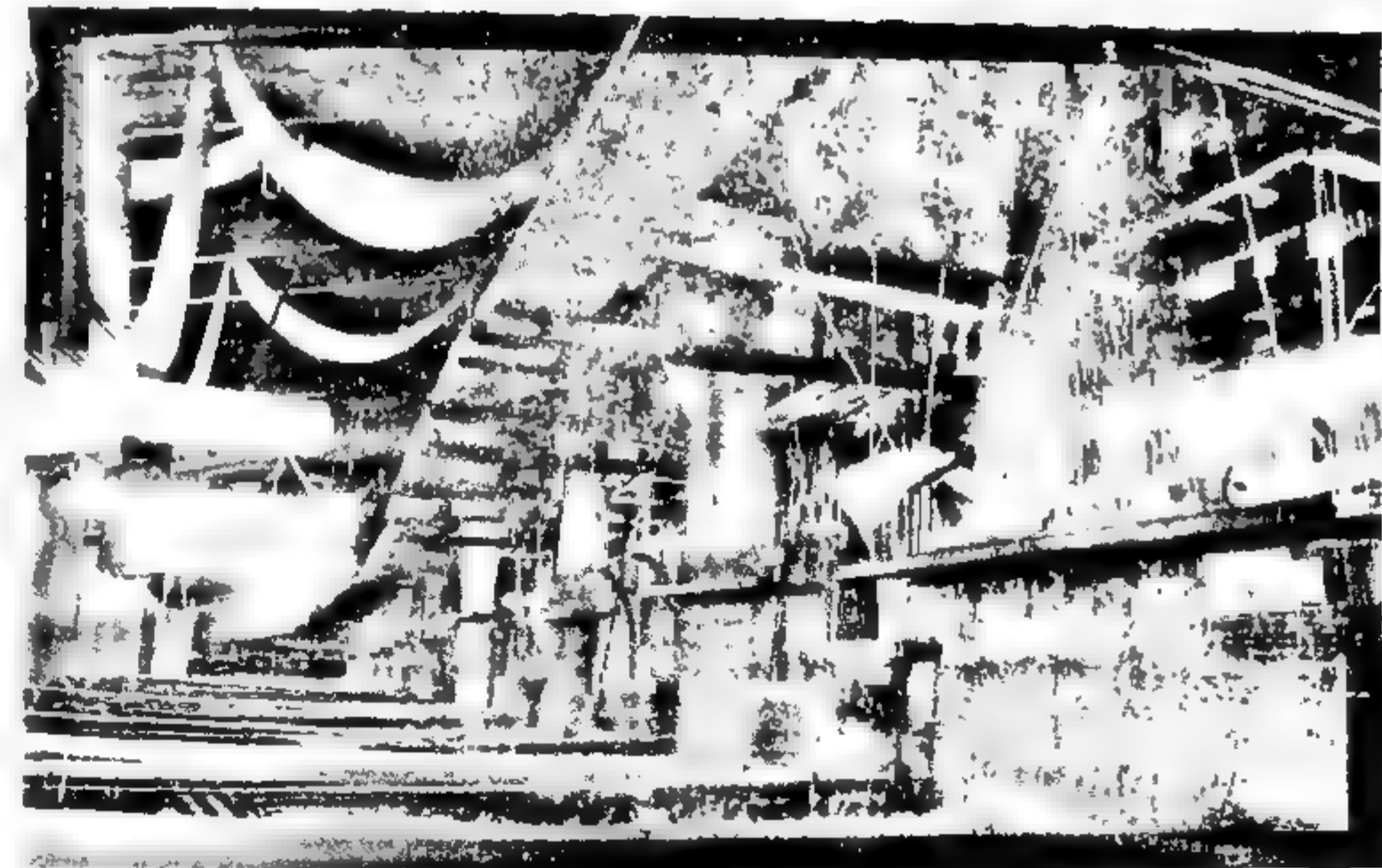
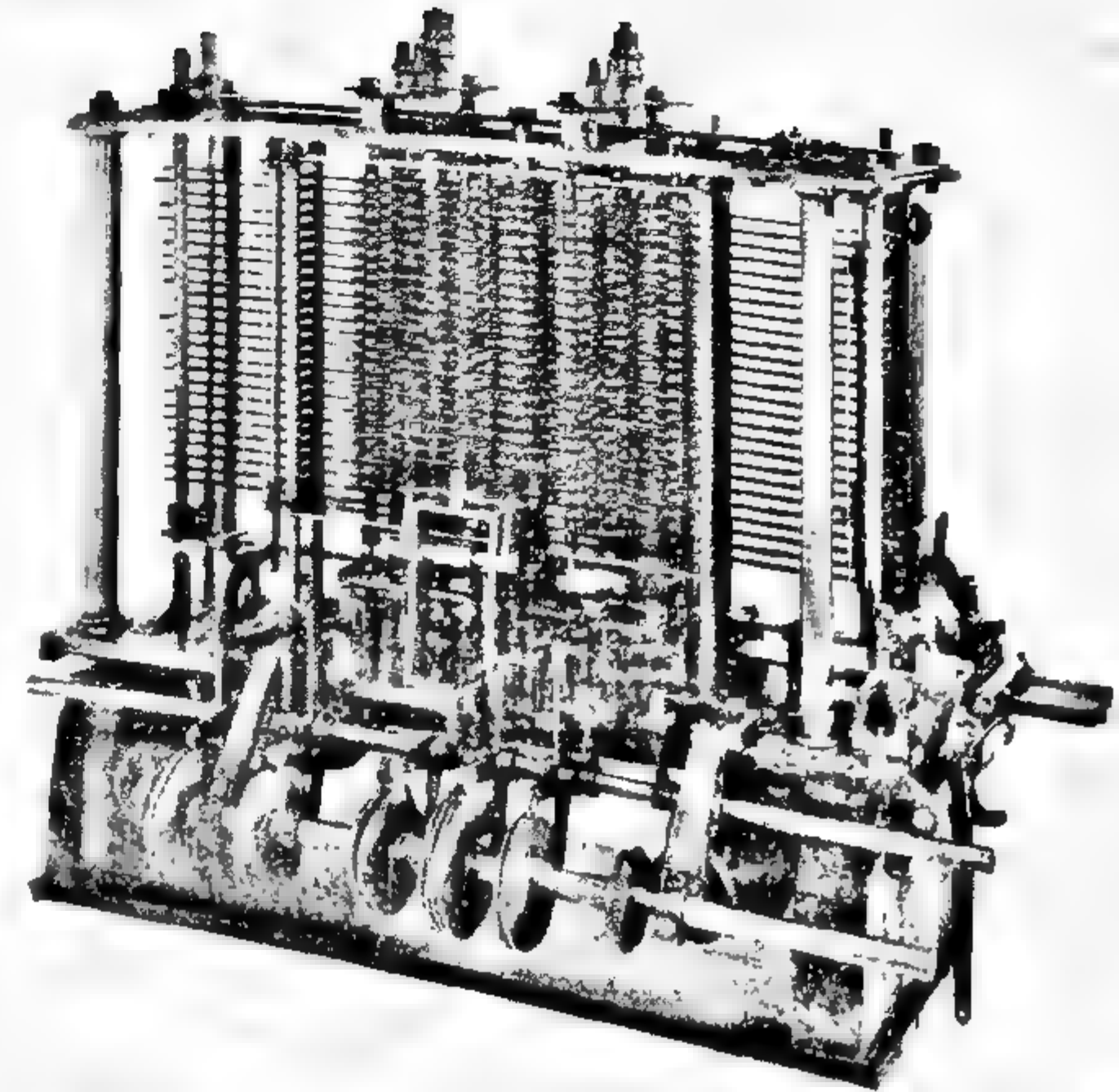


llaron la máquina Logic Theorist, que era capaz de demostrar teoremas nada triviales de lógica matemática. Esta innovadora tecnología se basó en todos los avances propuestos por Boole y De Morgan.

Un gran paso para la informática: las visionarias máquinas de Babbage

El inglés Charles Babbage (1791-1871) fue, además de matemático, un gran científico de la computación. Durante muchos años intentó llevar a cabo la construcción de la llamada máquina diferencial, una calculadora descomunal que tuvo que abandonar por su enorme complejidad tecnológica, inasequible para la época: constaba de 25 000 piezas, pesaba 13 600 kg y medía 2,4 m de altura. Era un gasto inasumible y el Gobierno británico dejó de financiarlo en 1832. Más tarde, entre 1833 y 1844, ideó una máquina aún más compleja que la anterior. Programable y capaz de realizar cualquier tipo de cálculo, la máquina analítica (fotografía superior de la página contigua) se considera el primer computador de la historia. Funcionaba con un motor a vapor y hacía uso de tarjetas perforadas como método de entrada. Para diseñarlo, Babbage se inspiró en el telar del tejedor y comerciante francés Joseph Marie Jacquard, célebre por ser el creador del primer telar programable (fotografía inferior de la página contigua), que operaba con ese tipo de tarjetas gracias a las cuales, de forma mecánica, se entrelazaban los hilos formando dibujos. La máquina de Babbage tenía, además, un procesador para calcular operaciones aritméticas, una unidad de control que determinaba qué tarea debía ser realizada, un mecanismo de salida y una memoria donde los números podían ser almacenados hasta ser procesados. En definitiva, un esquema muy similar al de un computador moderno.

Pero, a pesar de ser una máquina totalmente visionaria (que hubiera medido 30 m de altura por 40 de anchura), Babbage tampoco pudo llevarla a cabo por múltiples problemas financieros y de ingeniería. Pero años más tarde, la matemática Ada Love-



Arriba, máquina analítica de Charles Babbage que se conserva en el Science Museum of London. Abajo, telares Jacquard basados en el sistema de tarjetas perforadas, en el que se inspiró Babbage para dar entrada a los datos y programas con los que operaba su máquina.

Tan pronto como exista una máquina analítica, esta necesariamente guiará el curso futuro de la ciencia.

CHARLES BABPAGE

lace (1815-1852), que había conocido a Babbage en una fiesta social cuando apenas tenía diecisiete años, estudió a fondo los entresijos del aparato tras traducir al inglés lo que había escrito el militar y científico italiano Luigi Menabrea, considerado el autor del primer artículo científico de la historia de la informática, sobre la máquina analítica de Babbage. Lovelace se interesó en ella de tal manera que creó un programa ex profeso para solventar diversos problemas matemáticos, como un algoritmo para calcular los números de Bernoulli. Así fue como Lovelace, hija del poeta lord Byron, se convirtió

en la primera programadora de la historia, y ella y Babbage, en el primer equipo de ingenieros informáticos. Lovelace, quien fue además una de las primeras mujeres que entró en los anales de la ciencia y la tecnología, fue la encargada de la parte *soft* (software) y Babbage, de la parte *hard* (hardware). Dos tareas igual de complejas, a la par que inseparables. El lenguaje de programación ADA, desarrollado por el Departamento de Defensa de Estados Unidos, lleva este nombre en honor a Lovelace y sus contribuciones. El ADA ha sido durante muchos años la base para otros lenguajes de programación como el C++ o el Java. Aunque en su momento a Lovelace solo se le reconoció su papel como transcriptor de las notas de Babbage, investigaciones recientes muestran la originalidad de su trabajo de programación de la máquina analítica.

Visiones de futuro para recibir el siglo xx

Alan Turing todavía no había nacido cuando, en España, un célebre ingeniero, matemático e inventor español aportaba notables avances técnicos tanto en el área civil como en la militar e iba allanando el camino a los futuros genios de la computación. Es el caso del ingeniero de caminos, matemático e inventor cántabro Leonardo Torres Quevedo (1852-1936), quien, entre muchas

otras cosas, es considerado el heredero tecnológico de Charles Babbage. Fue él quien construyó las máquinas de cálculo que el inglés describió en un plano teórico. Torres Quevedo se inició en la carrera de las máquinas de cálculo empezando por las analógicas, que tratan de resolver problemas matemáticos, es decir, ecuaciones, mediante su traslado a fenómenos físicos. Los números y las variables se representan mediante magnitudes físicas, como las rotaciones de determinados ejes. Mediante estos ingeniosos aparatos, Torres Quevedo llegó a construir máquinas analógicas que permitían expresar mecánicamente la función $y = \log(10^{x+1})$, resolver ecuaciones de segundo grado con coeficientes complejos o, incluso, una máquina que permitía integrar funciones.

En el contexto de las máquinas analógicas, Torres Quevedo construyó El Ajedrecista, considerado el primer videojuego de la historia. A diferencia de otros artilugios como El Turco, El Ajedrecista jugaba automáticamente un final de rey y torre contra el rey de un oponente humano mediante un sistema de electroimanes. La máquina lograba la victoria en todas las ocasiones, aunque no fuera con el número mínimo de movimientos. Huelga decir que El Ajedrecista agrandó la fama mundial de Torres Quevedo, proporcionándole extensas menciones en prestigiosas revistas, como *Scientific American* («Torres and His Remarkable Automatic Device»).

De las máquinas analógicas, gobernadas por magnitudes físicas y tangibles, Torres Quevedo pasó a las máquinas electromecánicas, ahora sí, gobernadas ya por la electricidad, casi como un computador moderno. Torres Quevedo desarrolló la disciplina de la automática para servirse de ella y poder construir su primer computador, bautizado como «el aritmómetro», una máquina electromecánica capaz de realizar cálculos de forma autónoma con un dispositivo de entrada de comandos —una máquina de escribir—, una unidad de procesamiento y registros de valores —un sistema de poleas, agujas, escobillas, electroimanes y conmutadores—, y un dispositivo de salida —de nuevo, una máquina de escribir—. De hecho, estos bloques básicos son los que hoy en día componen cualquier computador, desde

EL TURCO, UN FALSO AUTÓMATA

El Turco fue un famoso fraude basado en un autómata que se decía que jugaba al ajedrez mediante un presunto sistema de inteligencia artificial. Construido por el inventor húngaro y excelente ajedrecista Wolfgang von Kempelen en 1769, era una cabina de madera de 110 x 60 x 75 cm, en la que había un maniquí vestido con un traje tradicional turco sentado en la parte posterior. La cabina tenía puertas que mostraban un mecanismo de relojería y, cuando se activaba el autómata, era capaz de competir al ajedrez contra un jugador humano experto. Sin embargo, la cabina era una mera ilusión óptica, ya que escondido en su interior había un maestro del ajedrez real que operaba el maniquí. Mediante un sistema de espejos se transmitía la visión del tablero de juego al maestro agazapado dentro del habitáculo. Los oponentes solían ponerse nerviosos al enfrentarse a un «revolucionario» sistema de inteligencia artificial y El Turco ganaba casi todas las partidas disputadas.

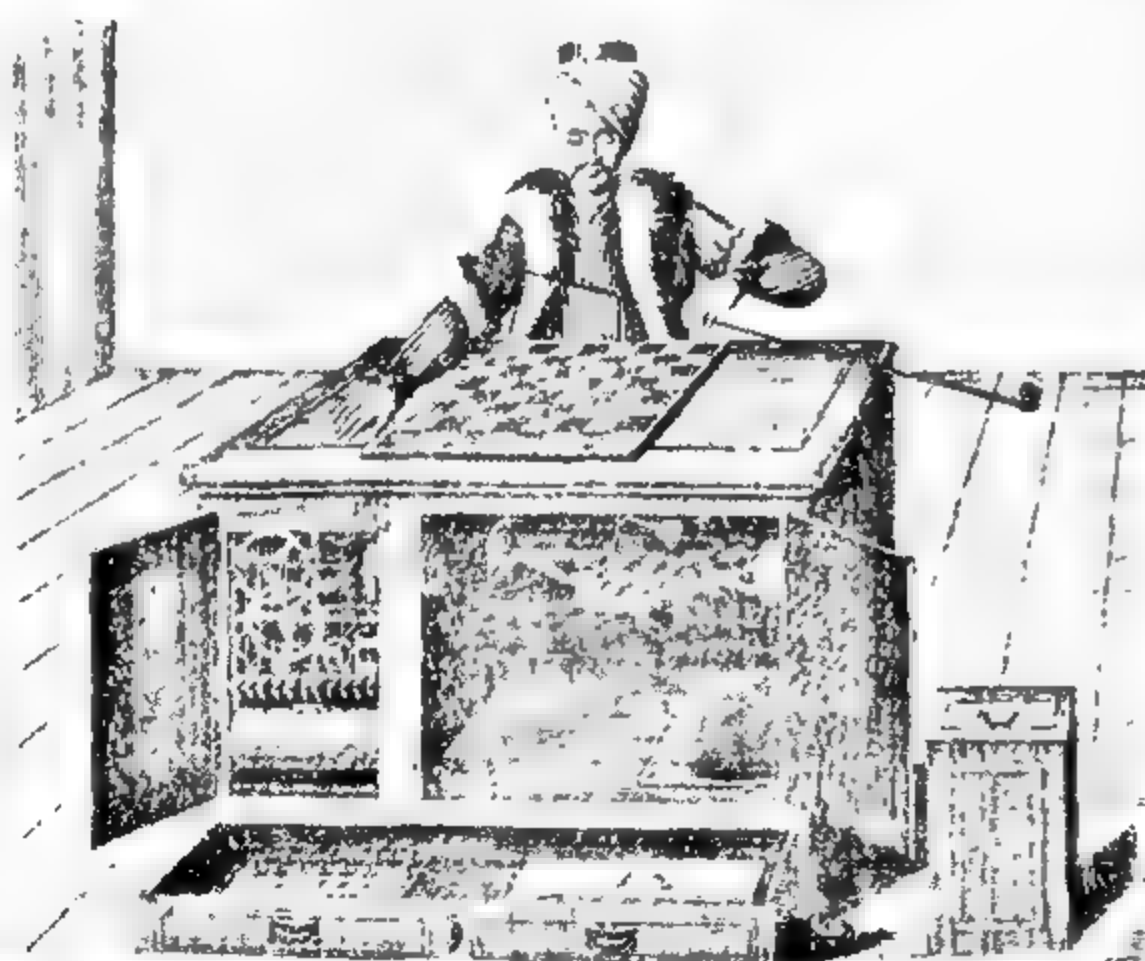
Fin de El Turco y revelación del secreto

Tras la muerte de Kempelen en 1804, El Turco pasó por muchas manos y continuó de gira, llegando a jugar contra adversarios tan célebres como Napoleón, en 1809, e incluso el mismo Babbage en 1820. En 1854, 85 años después de su construcción, El Turco fue destruido en un incendio que tuvo lugar en el museo de Filadelfia donde había sido donado, sin que su secreto hubiera sido revelado al público. Pocos años después, Silas Mitchell, hijo de su último dueño, desveló el misterio en una serie de artículos aparecidos en la publicación *The Chess Monthly*.

La versión del siglo XXI

El artificio ha inspirado, a *posteriori*, alguna iniciativa empresarial, como es el caso de un servicio que la empresa Amazon creó en 2005, llamado *Amazon Mechanical Turk*. Se trata de un servicio de externalización abierta de tareas o *crowdsourcing* (algo así como «fuentes surgidas de la multitud»), donde empresas de todo tipo buscan mano de obra barata para realizar trabajos normalmente repetitivos y de baja cualificación. Por ejemplo, seleccionar de entre 10000 fotografías todas aquellas en las que aparezcan caras humanas, o localizar personas desaparecidas mediante la visualización de millones de imágenes de satélite.

En este grabado de 1783 se muestra el mecanismo del célebre falso autómata jugador de ajedrez.



cualquier portátil hasta el mayor supercomputador del mundo. El mérito, por supuesto, también fue, en parte, de Charles Babbage, pero fue Torres Quevedo quien lo pudo construir y, lo más importante: ¡funcionó!

Quedaron así establecidas las bases preliminares de la inteligencia artificial, describiendo las máquinas como los elementos que deben desempeñar aquellas tareas para las cuales no es necesario pensar. De hecho, en una entrevista realizada en 1915 por *Scientific American*, el cántabro afirmó que casi todas las operaciones pueden ser realizadas por una máquina, incluso aquellas que requieren la participación de una gran capacidad intelectual. En uno de sus ensayos, como veremos, se adelantó incluso al célebre Turing discutiendo sobre el experimento bautizado como «test de Turing», y sus implicaciones filosóficas.

Rebatiendo a Descartes, quien afirmó que un autómata jamás sería capaz de mantener un diálogo razonable, Torres Quevedo dijo:

No hay entre los dos casos la diferencia que veía Descartes. Pensó sin duda que el autómata, para responder razonablemente, tendría necesidad de hacer él mismo un razonamiento, mientras que en este caso, como en todos los otros, sería su constructor quien pensara por él de antemano. Creo haber mostrado, con todo lo que precede, que se puede concebir fácilmente para un autómata la posibilidad teórica de determinar su acción en un momento dado, pesando todas las circunstancias que debe tomar en consideración para realizar el trabajo que se le ha encomendado.

Además de todas sus aportaciones a la inteligencia artificial, este Leonardo nacido en pleno siglo XIX diseñó un nuevo tipo de dirigibles que fueron utilizados en la Primera Guerra Mundial con fines de vigilancia naval. Más tarde, proyectó una versión

Hay una clase de autómatas que ofrecen el mayor interés: los que imitan, no los gestos, sino las acciones del hombre.

LEONARDO TORRES QUEVEDO

mejorada con la intención de convertirla en la primera aeronave capaz de cruzar el Atlántico. Sin embargo, por problemas de financiación, el proyecto se retrasó y fueron dos británicos, Arthur Brown y John Alcock, los que lo consiguieron con un biplano en 1919. Como ingeniero, también trabajó en el diseño de nuevas formas de teleféricos, funiculares y transbordadores, siendo el más famoso de estos el Spanish Aerocar, que cruza las cataratas del Niágara.

Pero eso no fue todo. En el ámbito de la automática, fue responsable del desarrollo del primer elemento teledirigido de la historia, un autómata llamado Telekino que ejecutaba órdenes retransmitidas mediante ondas hercianas, igual como funciona hoy en día cualquier juguete teledirigido o el mando a distancia del televisor. Siempre trabajó con una óptica puramente comercial y de mercado, patentando todo lo que podía y obteniendo rendimientos de sus invenciones. En el caso de los dirigibles que fueron usados en la Primera Guerra Mundial, fueron fabricados por la compañía francesa Astra, la cual compró la patente al propio inventor.

CAPÍTULO 2

////////////////////////////////////

Con Turing se hizo la luz

Alan Turing está considerado el padre de la inteligencia artificial. Esta suerte de filósofo, matemático e ingeniero no solo marcó un antes y un después en esta disciplina, sino que también estableció la línea divisoria entre el pasado y el futuro de la computación.

Sin duda alguna, las aportaciones a la informática de Alan Turing (1912-1954) fueron trascendentales. Aportó grandes avances en cuatro áreas de aquella informática primitiva: la teoría de la computación, la arquitectura de computadores, el criptoanálisis y, por supuesto, la inteligencia artificial. Lo importante es que estas aportaciones estuvieron muy relacionadas entre sí: el desarrollo de la teoría de la computación permitió idear el concepto de máquina de Turing y, en base a ello, se pudo perfeccionar la arquitectura de los computadores propuestos por Babbage e implementados por Torres Quevedo. Gracias a estos nuevos ordenadores y a su potencia de cálculo, el científico británico avanzó enormemente en el campo del criptoanálisis, siendo él uno de los principales artífices en la ruptura de los códigos de la máquina Enigma, que los nazis emplearon durante la Segunda Guerra Mundial para el cifrado y descifrado de mensajes; este hecho facilitó en gran medida la victoria de los aliados en la contienda. Una vez acabada la guerra, con la potencia de cálculo ganada y solucionados los temas más candentes que planteaba el conflicto armado, Turing se volcó en el ámbito de la inteligencia artificial.

EL TEST DE TURING O CÓMO DETERMINAR SI UNA MÁQUINA ES INTELIGENTE

Fue este genial y multifacético científico quien, por primera vez y de forma práctica, respondió a la pregunta de si una determinada entidad no humana es o no inteligente. En 1950 propuso un test basado en una idea muy simple: si una máquina se comporta en todos los aspectos como un ente inteligente, entonces es que debe de ser inteligente.

Pero dejemos por un momento a un lado al destacado matemático, criptógrafo y lógico y demos un pequeño paseo por el terreno de la ciencia ficción. Y es que hay un diálogo magistral en la película *2001: Una odisea del espacio* dirigida por Stanley Kubrick y estrenada en 1968, que ilustra a la perfección el razonamiento que tuvo Turing en la definición de inteligencia. En este diálogo, un periodista de la BBC entrevista al equipo de astronautas y al supercomputador HAL 9000, a cargo de las funciones vitales de la nave espacial *Discovery*. Después de varias preguntas y respuestas sobre HAL 9000, el astronauta Dave Bowman concluye que la máquina actúa como si tuviera emociones genuinas, pero que si las tiene o no es algo que nadie puede contestar. Por tanto, como decía Turing, si HAL 9000 se comporta en todos los aspectos como un ente inteligente, entonces es que debe de serlo.

En dicha entrevista se habla de conceptos tan trascendentes como la consciencia o las emociones genuinas, así como del orgullo o el divertimento. ¿Puede una máquina divertirse? HAL 9000 cuenta que se divierte trabajando con los astronautas y estos afirman que uno se hace rápidamente a la idea de que HAL 9000 es uno más. ¿Es esto un síntoma de inteligencia artificial? Según Turing, desde luego que sí.

Pero veamos exactamente en qué consiste el test de Turing. El test, que no debe entenderse en sentido literal sino metafórico, sitúa a un evaluador humano y a la máquina sujeta a la valoración en dos salas separadas por un tabique que imposibilita la visión directa (figura 1). Entre ambos existe un teclado y una pantalla que usarán para comunicarse. A continuación, el evaluador

HAL 9000: LA ENTREVISTA

El supercomputador de a bordo de la nave espacial que aparece en la película *2001: Una odisea del espacio* se llama HAL 9000 (por el acrónimo en inglés *Heuristically Programmed ALgorithmic Computer*) y ostenta una inteligencia artificial que le hace razonar de forma parecida a los humanos. Así responde a las preguntas de un supuesto periodista.



Ojo-cámara con el que se representó al supercomputador HAL 9000.

Entrevistador: Buenas tardes HAL. ¿Cómo va todo?

HAL 9000: Buenas tardes Mr. Amor. Todo va extremadamente bien.

Entrevistador: HAL, tú tienes una responsabilidad enorme en esta misión. En cierta manera, la responsabilidad más grande de todos los elementos de la misión. Tú eres el cerebro y el sistema nervioso de la nave y tu responsabilidad incluye estar atento a los hombres en hibernación. ¿Esto te ha causado alguna vez alguna falta de confianza?

HAL 9000: Déjeme exponerlo de la siguiente forma. La Serie 9000 es el computador más fiable jamás construido. El computador 9000 no ha cometido nunca ningún fallo ni distorsionado ningún dato. Somos todos, de forma literal, a prueba de tontos, e infalibles.

Entrevistador: HAL, a pesar de tu enorme intelecto, ¿te has sentido frustrado alguna vez por tu dependencia de la gente para ejecutar tus acciones?

HAL 9000: No, en lo más mínimo. Me divierte trabajar con gente. Tengo una relación estimulante con el Dr. Poole y el Dr. Bowman. Mis responsabilidades en la misión se reparten sobre todo en el funcionamiento de la nave, así que estoy constantemente ocupado. Estoy dando lo máximo de mí, lo cual creo que es todo a lo que puede aspirar cualquier entidad consciente.

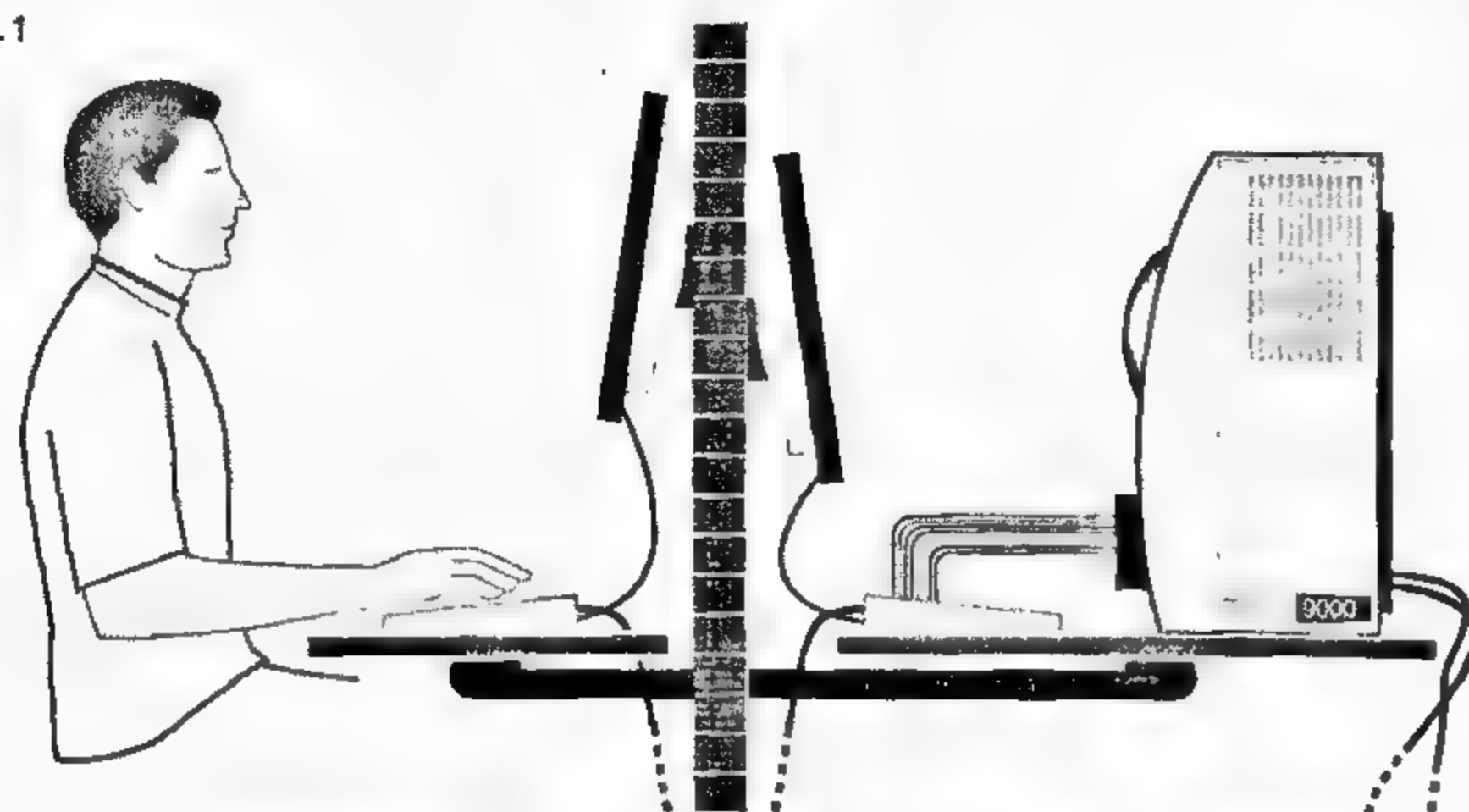
Entrevistador: Dr. Poole, ¿cómo es vivir durante la mayor parte del año de forma tan estrecha con HAL?

Dr. Poole: Bueno, es muy parecido a lo que ha dicho usted antes sobre él. Él es como un sexto miembro de la tripulación. Uno se ajusta muy rápidamente a la idea de que él habla, y piensas de él que realmente es otra persona.

Entrevistador: Al hablar con el computador, uno tiene la sensación de que él es capaz de dar respuestas emocionales. Por ejemplo, cuando le pregunté acerca de sus habilidades, percibí un cierto orgullo en su respuesta acerca de su exactitud y perfección. ¿Ustedes creen que HAL tiene emociones genuinas?

Dr. Bowman: Bueno, él actúa como si tuviera emociones genuinas. Por supuesto que está programado de esa manera para que sea más fácil para nosotros hablar con él, pero en cuanto a si tiene o no tiene sentimientos reales, es algo que no creo que nadie pueda contestar.

FIG. 1



X realiza las preguntas

Y responde a las preguntas

Esquema del test de Turing. Nótese que no existe contacto visual. El test es superado si X no detecta si Y es un humano o una máquina.

puede hacer preguntas a la máquina y esta puede responder. Si el humano considera que quien está respondiendo es una persona, entonces la máquina evaluada es inteligente y, por tanto, se considera que posee inteligencia artificial.

Cuando se menciona que el test de Turing debe entenderse en sentido metafórico significa que no es necesario que ese sistema presuntamente inteligente sea capaz de entender las preguntas y de responder mediante un lenguaje natural, sino que el humano evaluador debe poder afirmar que la respuesta obtenida la ha planteado otro humano. Veamos, por ejemplo, el resultado de otro de los legados de Turing: un programa de ordenador que era capaz de escribir cartas de amor como esta:

*Querido encanto,
Eres mi ávido sentimiento amigo.
Mi afecto se asocia extrañamente a tu deseo pasional.
Mi deseo ansía tu corazón.*

*Eres mi soñadora compasión: mi tierno deseo.
Hermosamente tuyo.*

M.U.C. (Computador de la Manchester University)

Si recibiéramos una carta de amor como esta, ¿el programa de ordenador usado estaría pasando el celeberrimo test? Huelga decir que el nombre de test de Turing fue acuñado con posterioridad. En realidad, en su artículo original de 1950, Turing lo bautizó como «el juego de imitación», inspirándose en un popular juego inglés de época victoriana que consistía en encerrar en dos habitaciones separadas a un hombre y a una mujer. Entonces un tercer sujeto tenía que adivinar dónde estaba la mujer y dónde el hombre, y podía formular cuantas preguntas quisiera a las dos personas que había en las habitaciones. El juego tenía una particularidad: la mujer siempre estaba obligada a decir la verdad, y el hombre, a mentir.

La variante del juego de la imitación propuesto por Turing consistía en intercambiar a la persona encerrada en la habitación por una máquina. Si el interlocutor confunde la máquina con la persona, entonces podemos afirmar que estamos ante una máquina inteligente. En caso contrario, no se puede concluir nada. Es decir que, si la máquina no pasa el test de Turing, no se puede afirmar que no sea inteligente.

CRÍTICAS DE ÍNDOLE FILOSÓFICA CUESTIONAN EL TEST

El test de Turing, sin embargo, no estuvo exento de polémica. Cuando el científico formuló el procedimiento para evaluar la supuesta inteligencia de cualquier entidad, recibió un aluvión de críticas por parte de algunos teóricos. ¿Se puede decir que una máquina es inteligente si simplemente responde a las preguntas basándose en un diccionario enorme de preguntas y respuestas? ¿O la inteligencia consiste en algo más que cierto comportamiento aparente como, por ejemplo, la existencia de consciencia?

La principal crítica del test de Turing la escenificó el filósofo estadounidense John Searle (n. 1932) mediante el experimento denominado «la habitación china». Imaginémosnos que cualquiera de nosotros, que no sabe ni una palabra de chino, entra en una habitación cerrada en cualquier lugar de Pekín. A

Supuestamente el cerebro humano es algo parecido a una libreta que se adquiere en la papelería: muy poco mecanismo y muchas hojas en blanco.

ALAN TURING

continúa, se invita a que las personas que pasan por allí introduzcan preguntas escritas a las personas que están dentro de la habitación a través de una abertura. Estas tienen a su alcance un manual donde están representados los caracteres chinos que conforman las respuestas a cada una de las posibles preguntas que se pueden realizar. Pongamos por ejemplo que alguien escribe en chino «¿Hace frío ahí dentro?» y pasa la cuestión por la rendija. Acto seguido, la persona que hay dentro de la habitación lee los caracteres chinos, busca en el manual una posible respuesta a esta pregunta y la transcribe, trazo por trazo, en una nueva hoja que entrega al interlocutor. La respuesta dice «No, hace mucho calor». Lógicamente, quien ha hecho la pregunta habrá recibido una contestación coherente en su idioma y pensará que quien haya dentro de la habitación conoce el chino perfectamente. Pero, como sabemos, el de dentro de la habitación ha basado todas sus respuestas en un manual de correspondencias, sin entender ni una sola palabra de chino.

Entonces ¿podemos sospechar que una máquina que supera el test de Turing puede estar engañándonos de una manera similar a como lo hace la habitación china? La respuesta es que no: la habitación china es un escenario engañoso pues, en realidad, si bien es cierto que la persona que está dentro de ella no sabe chino, quien realmente contesta las preguntas es el conjunto de entidades formado por la persona y el manual. Aunque no se puede decir que el manual «sepa» chino, en todo caso, sí ha sido confeccionado por alguien que, efectivamente, debe conocer el idioma.

Desde una perspectiva práctica, hoy en día se considera que una nueva tecnología es inteligente si es capaz de resolver un problema de una manera creativa, algo que desde siempre se ha considerado potestad exclusiva del cerebro humano. Un ejemplo representativo de una tecnología que parece inteligente pero que no se considera como tal es la de los primeros *sistemas expertos* que aparecieron en la década de 1970. Un sistema experto es un programa informático en el que se han implementado unas determinadas reglas, más o menos complejas, y el problema puede actuar de manera autónoma en el control de ciertos sistemas. Un ejemplo puede ser un programa de ordenador con una enorme lista de síntomas médicos programados que, dado un nuevo paciente y sus síntomas, puede determinar el tratamiento médico que el paciente necesita. Sin embargo, si el sistema no es capaz de crear una nueva regla deducida de las anteriores ni inventar un nuevo tratamiento cuando la situación lo requiera, se considera que no es creativo y, por ende, que no es inteligente.

Aparte de la crítica de la habitación china, otros pensadores contemporáneos, como los expertos en inteligencia artificial Rolf Pfeifer, del Departamento de Informática de la Universidad de Zúrich, en Suiza, y Christian Scheier, del Instituto Tecnológico de California, en Estados Unidos, han lanzado otras críticas. En particular, han argumentado en favor del denominado *problema del marco de referencia*. Esta argumentación apunta que no podemos concluir la complejidad interna de una máquina a partir de su comportamiento. También ha habido otras críticas en el plano teológico, vinculadas a la idea de algunos pensadores que opinan que pensar es una facultad del alma inmortal humana. Desde el ámbito de la creatividad se ha afirmado que los ordenadores son incapaces de tener originalidad —un concepto conocido como la *objeción de lady Lovelace*, porque ella afirmó que las máquinas no pueden aprender de forma independiente— o la *objeción matemática*, que afirma que a través del teorema de la incompletitud de Gödel se puede demostrar que existen ciertos límites a los problemas a los que un computador puede dar respuesta.

CÓMO PASAR EL TEST DE TURING SIENDO UNA MÁQUINA

Hoy en día existen multitud de ejemplos que podrían pasar el test de Turing. Ya hemos visto el ejemplo de *2001: Una odisea del espacio*, aunque esto es pura ciencia ficción. También ha habido recientes hitos de la inteligencia artificial que pasan, de largo, por el test de Turing, como el realizado por el supercomputador de IBM llamado Deep Blue, que venció al campeón de ajedrez Kaspárov, o el sistema informático Watson, también desarrollado por IBM, que ganó en el popular juego televisivo americano *Jeopardy!* Pero dejemos estos dos ejemplos recientes para más adelante y veamos ahora algunos ejemplos «clásicos», como ELIZA o PARRY. Ambos son sistemas del tipo bot de charla o conversacional, programas que simulan tener una conversación mediante respuestas automáticas, de finales de los años sesenta y principios de los setenta respectivamente, e imitan una conversación con una persona. ELIZA trata de hablar como una persona normal, mientras que PARRY imita a una persona paranoica. Veamos un ejemplo de conversación con ELIZA:

ELIZA: Hola, vamos a hablar.

Interlocutor: Hola, ¿cómo estás?

ELIZA: ¿Por qué estás interesado en saber si estoy o no estoy?

Interlocutor: Hoy es un gran día.

ELIZA: Ya veo.

Interlocutor: Me tomaría una cerveza.

ELIZA: Eso es bastante interesante.

Interlocutor: ¿A ti te gusta la cerveza?

ELIZA: Hablemos de ti, no de mí.

Interlocutor: Vale.

ELIZA: Bueno, vale, me gusta la cerveza. Dime más.

Interlocutor: Me tengo que ir. Adiós.

ELIZA: Dime más.

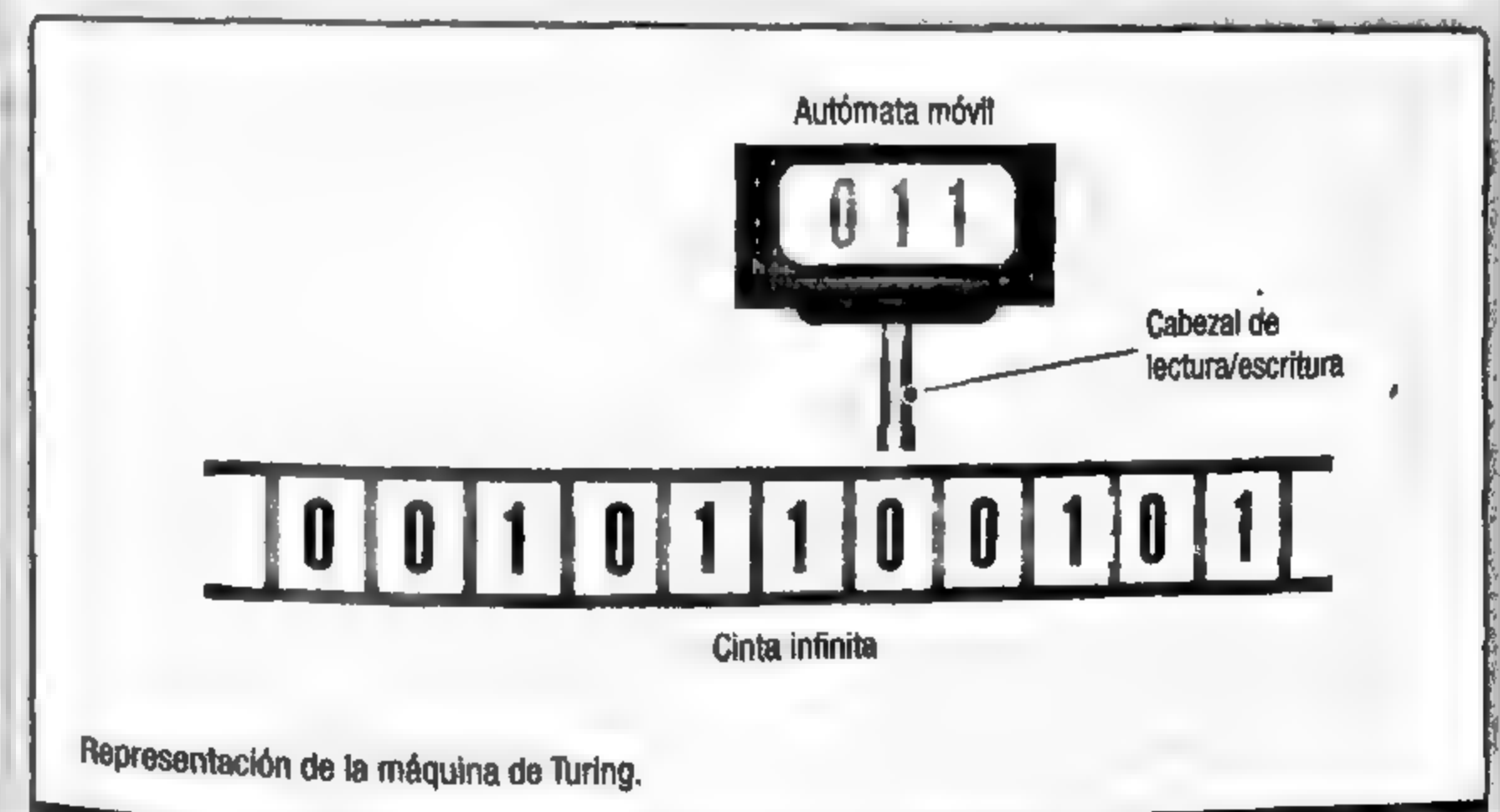
Como puede observar el lector, el diálogo es un poco pobre, pero esto es todo lo que se podía conseguir en los años sesenta.

LA MÁQUINA DE TURING

La máquina de Turing es un modelo teórico de computador universal, es decir, un sistema capaz de procesar cualquier dato de entrada y calcular una salida en un tiempo finito. En el plano teórico, una máquina de Turing consta de una cinta infinita con símbolos escritos sobre ella, de un cabezal que puede desplazarse a izquierda o derecha sobre la cinta, leer los símbolos, borrarlos y escribir otros nuevos sobre la cinta, así como de unas reglas que determinan el comportamiento del cabezal para cada posible símbolo que se detecte sobre la cinta. En la informática práctica estas reglas representarían un programa informático, y la cinta, el sistema de entrada/salida del programa y un registro sobre su estado de ejecución.

Lenguajes de programación Turing-compatibles

Hoy en día, cuando se desarrolla un nuevo lenguaje de programación, como podría ser C++, Python, Java, etc., lo primero que se debe demostrar de manera formal, es decir, en términos de demostración matemática, es que el nuevo lenguaje es Turing-compatible, o, en otras palabras, que el nuevo lenguaje es equivalente a una máquina de Turing. Esto es importante porque cuando se diseña un nuevo lenguaje de programación se pretende que este sea capaz de resolver, al menos desde un punto de vista teórico, cualquier problema. La manera más sencilla de demostrar si un nuevo lenguaje es Turing-compatible es programando una máquina de Turing utilizando el nuevo lenguaje. Si efectivamente la máquina se puede implementar, habremos demostrado que es capaz de implementar un computador universal y, por tanto, que el lenguaje de programación utilizado es, a su vez, un computador universal capaz de resolver cualquier problema que emane del ingenio de su programador.



Veamos un divertido diálogo de besugos entre ELIZA y PERRY que se puede encontrar publicado en internet:

ELIZA: Buenas noches. Háblame sobre tus problemas.

PERRY: A veces la gente me saca de quicio.

ELIZA: No estoy segura de si te acabo de entender.

PERRY: Deberías prestar más atención.

ELIZA: Supón que te presto más atención.

PERRY: Estás sujeta a tu propia opinión.

ELIZA: ¿Qué te hace suponer que estoy sujeta a mi propia opinión?

PERRY: Me importa estar alrededor de la gente.

ELIZA: Por favor, continúa.

PERRY: Me fui a la pista de Bay Meadows hace un rato.

ELIZA: ¿Y eso qué te sugiere?

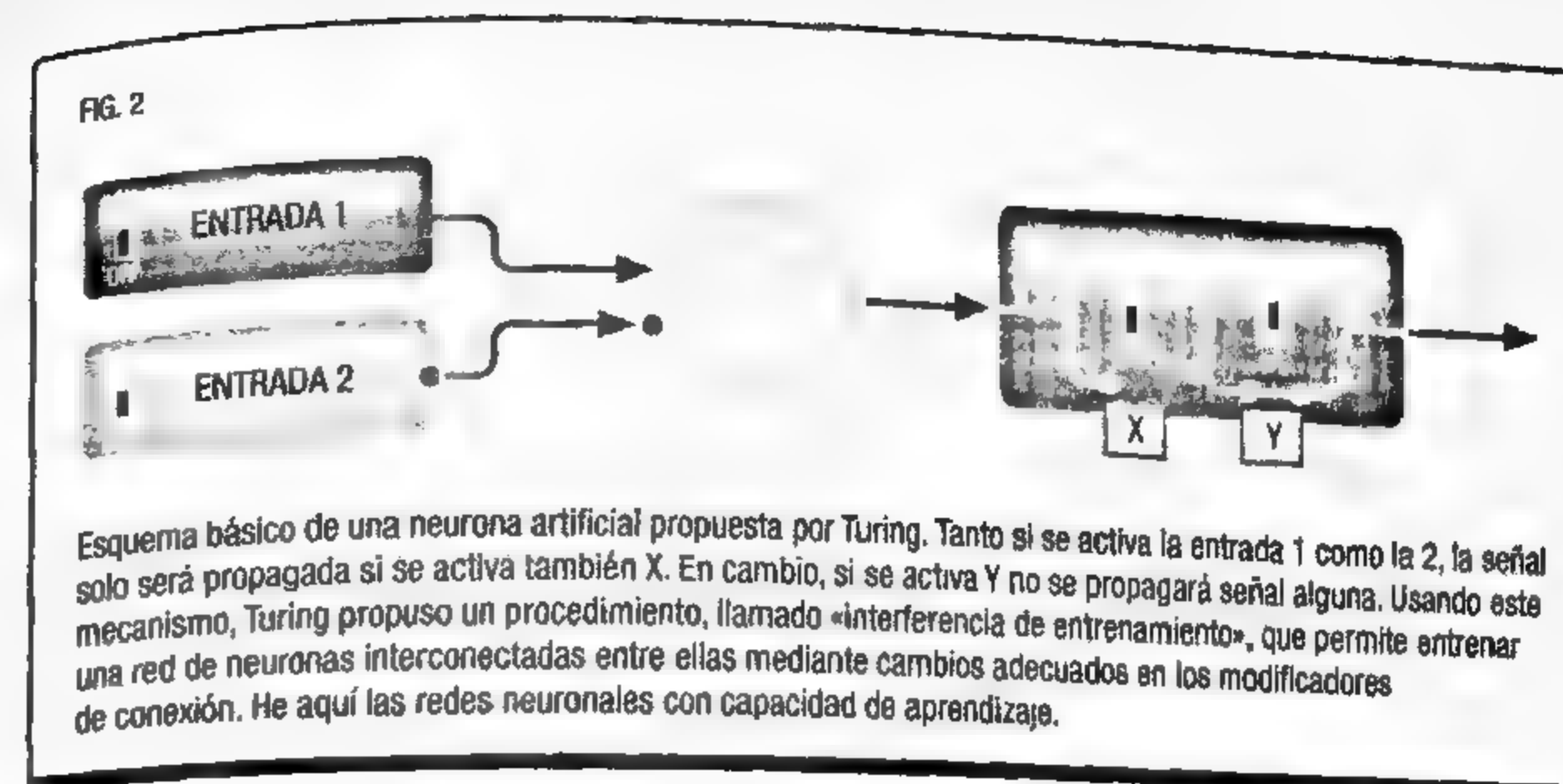
PERRY: Me fui a las carreras.

EL MODELO CONEXIONISTA: LAS NEURONAS ARTIFICIALES

Aunque la gran aportación de Turing en el campo de la inteligencia artificial fue la reflexión filosófica que le condujo a su famoso test, también le debemos los fundamentos de las redes neuronales, esas potentes herramientas compuestas por neuronas artificiales que han ido desarrollándose a lo largo de los años hasta convertirse en uno de los pilares centrales de la inteligencia artificial moderna.

Esa reflexión partió de la idea de que, en el momento del nacimiento de cualquier persona, el cerebro es una máquina desorganizada que, a medida que va madurando, se va estructurando a través del aprendizaje. A partir de aquí, Turing propuso lo que él denominó *máquina desorganizada de tipo B* o lo que modernamente se conoce como «neurona artificial» (figura 2), que, en contraposición a la de *tipo A*, puede ser entrenada mediante unas líneas de entrenamiento, llamadas *modificadores de conexión*.

Aunque en realidad Turing fue el primero en proponer este modelo, el mérito definitivo del asunto se lo llevaron los neuro-



científicos estadounidenses Warren McCulloch y Walter Pitts, al ser su modelo de neurona artificial, creado en 1943, el que realmente ha pasado a ser el fundamento de las redes neuronales actuales.

EL EMPIRISMO RADICAL: WITTGENSTEIN CONTRA TURING

El periodo de la historia de la ciencia previo a Turing es realmente interesante. Desde que Darwin publicó su obra *El origen de las especies*, en 1859, hasta la Segunda Guerra Mundial, los pensadores de la época —incluyendo científicos, filósofos y teólogos— no hacían más que enfrentar continuos cambios de paradigma. Si Darwin vino a demostrar que el hombre no es fruto de ningún diseño inteligente, sino el resultado de una mera evolución desde los primates, Sigmund Freud estableció que nuestras fobias y filias no son actos voluntarios, sino fruto de algo llamado «subconsciente» que guía subrepticamente nuestro devenir en contra, muchas veces, del propio raciocinio. Acto seguido, con la llegada de la física cuántica y la relatividad, se determinó que la materia, es decir, las partículas, presentan características tanto corpusculares como ondulatorias. Y, finalmente, Edwin Hubble enseñó al mundo que tanto el Sol como nuestro planeta

LA MANZANA DE TURING

A lo largo de su vida y a pesar de sus enormes contribuciones individuales a la victoria aliada en la Segunda Guerra Mundial, Turing fue perseguido por la justicia británica debido a su homosexualidad. En el transcurso de su juicio decidió no defenderse, al considerar que no tenía de qué disculparse. En la sentencia, Turing escogió la castración química antes que ir a prisión. El «tratamiento» consistía en una potente serie de inyecciones de hormonas, lo que le desencadenó el desarrollo de glándulas mamarias y una profunda depresión. A nivel científico, esta época de su vida coincidió con una etapa en la que estaba trabajando en lo que él denominó la «nueva mecánica cuántica», pero que, pese a su entusiasmo, no llegó a resultado alguno.

Finalmente, el 7 de junio de 1954, a los cuarenta y un años, se suicidó ingiriendo parte de una manzana envenenada con cianuro potásico. Una drástica decisión tomada, seguramente, debido a la humillación pública que sufrió y a los efectos del tratamiento hormonal y a la depresión que este le causó. A partir

de ahí nacieron dos «mitos» sobre Turing y las manzanas. Por una parte, cuentan que el suicidio de Turing se inspiró en la escena de la película de Disney *Blancanieves y los siete enanitos*, en la que la bruja envenena una manzana al sumergirla en una olla con una poción para entregársela a Blancanieves. La otra historia apunta que la manzana con la bandera multicolor de fondo que Apple usó como logotipo fue una suerte de homenaje a Turing, pero desde la compañía fundada por Steve Jobs eso se ha desmentido en reiteradas ocasiones. No fue hasta el año 2009, veinticinco años después de su lamentable muerte, cuando el primer ministro británico Gordon Brown pidió disculpas oficiales por el miserable trato que recibió este genio de la computación durante los últimos años de su vida.



Estatua de Alan Turing inaugurada el 23 de junio de 2001 en Sackville Park, Manchester. En la mano derecha sostiene una manzana, que podría aludir a su muerte, ya que supuestamente se suicidó al comer esta fruta envenenada con cianuro potásico.

son puntos insignificantes en la inmensidad del universo. De repente quedó claro que no controlamos nada: ni nuestra propia evolución, ni nuestra mente, ni la materia, ni nuestro lugar en un universo infinito en el que apenas somos una simple mota de polvo.

En ese marco tan interesante desde el punto de vista epistemológico, llegó Turing y empezó a plantear la inteligencia artificial. Mientras las bases de la filosofía se desmoronaban —las de la teología, por supuesto, ya lo estaban desde hacía tiempo—,

el Turing más filósofo comenzó a hablar de lo que «en teoría» tendría que pasar para que podamos decidir si una entidad es inteligente o no. Como es fácil de imaginar, esto no fue bien recibido en los círculos filosóficos «clásicos». A Turing le surgió un contrincante de talla, el filósofo austriaco Ludwig Wittgenstein, con el que mantuvo profundos desacuerdos intelectuales.

Acostumbramos a decir que Turing fue, ante todo, un matemático, aunque también, de forma menos relevante, fue un filósofo. Con Wittgenstein sucede justo a la inversa, es decir, primero fue un filósofo y en menor grado, matemático. Se dice que el primero fue alumno del segundo, pero que Turing dejó de ir a las clases de Wittgenstein por los profundos desacuerdos intelectuales que tenía con él. Probablemente, harto del reciente devenir de la ciencia, Wittgenstein inauguró, en su primera etapa filosófica, una nueva rama de la filosofía de la ciencia, el positivismo lógico, radicalmente empírica. Un empirismo radical que impide, en principio, elucubrar disquisiciones de elementos que están fuera del plano de la realidad, tal y como lo estaban las «máquinas pensantes» en aquella época. Pero quizá Wittgenstein también elaboró esta filosofía como réplica a las velocidades del tiempo cambiantes explicadas por Einstein en su teoría de la relatividad, a los gatos atrapados en cajas que no sabemos si están o no están —en alusión a la famosa metáfora de Schrödinger que trata de explicar la física cuántica— o a un ser humano que «simplemente» es una ligera mejora del mono.

La filosofía es una lucha contra el hechizo de nuestra inteligencia por el lenguaje.

LUDWIG WITTGENSTEIN

EL TRACTATUS LOGICO-PHILOSOPHICUS DE WITTGENSTEIN

Ludwig Wittgenstein (Viena, 1889—Cambridge, 1951) es uno de los grandes filósofos que nos ha dejado el siglo xx. En su trabajo se distinguen claramente dos etapas, y en la segunda, haciendo gala de una honestidad brutal, renegó de la primera. No en vano, a menudo le gustaba autoidentificarse con aquel que incendió la Biblioteca de Alejandría. Su primera etapa filosófica quedó plasmada en su única obra publicada en vida, el *Tractatus Logico-Philosophicus*, en la que estableció que la estructura profunda del lenguaje, es decir, la forma lógico-matemática que el lenguaje adopta, se corresponde con la de la realidad que nos envuelve. Por tanto, dicha realidad y el lenguaje adoptan una misma esencia lógica que permite que las palabras figuren las cosas, mientras que las proposiciones figuran los hechos.

La realidad (plausible) por encima de las matemáticas

Esto eliminaba de un plumazo la capacidad de filosofar —o dialogar— sobre cosas que no estén en el plano de la realidad, como cualquier deidad —y por ello Wittgenstein elimina la metafísica— o cualquier otro concepto en el plano de la ciencia ficción, como puede ser un concepto no desarrollado de inteligencia artificial contemporánea al pensador. En definitiva, según Wittgenstein, debatir sobre si un determinado test puede discernir si una máquina que aún no existe es inteligente o no, queda fuera del alcance del lenguaje. Del *Tractatus Logico-Philosophicus* destacan dos anécdotas. La primera es que la obra fue escrita en el frente, mientras Wittgenstein luchaba como voluntario en el ejército austriaco en la Primera Guerra Mundial, y la segunda es que el autor afirmaba que el libro se componía de dos partes: la escrita y la no escrita, siendo la segunda la más importante de las dos. Aunque parezca absurda, esta anécdota tiene una importancia caudal, ya que Wittgenstein se había propuesto delimitar completamente la disciplina sobre la que ensayaba y la mejor forma de delimitar un área de conocimiento es guardar silencio sobre lo que queda fuera de esta.

Donde dije inmutable, digo depende

La segunda etapa filosófica de Wittgenstein la caracteriza su obra *Investigaciones filosóficas*, en la que trabajó veinte años y que no fue publicada hasta dos años después de su muerte. En esta segunda etapa defendió vehementemente la idea de que el significado de una palabra no es algo inmutable. Es decir, que los significados de las palabras varían en función del uso que se haga de estas en cada contexto lingüístico. La influencia de Wittgenstein en la filosofía posterior ha sido muy relevante. De hecho, este impactó notablemente en la corriente filosófica del positivismo lógico, una rama de la filosofía de la ciencia que reduce el conocimiento a lo estrictamente empírico y verificable.

Ludwig Wittgenstein según el artista alemán Christiaan Dirk Tonnis. Antes que filósofo, matemático, lingüista y lógico, Wittgenstein fue ingeniero e incluso presentó una patente para un motor de reacción. Sin embargo, pronto se sintió subyugado por la filosofía de las matemáticas, siendo muy crítico tanto con sus colegas como con él mismo. Una de sus frases célebres reza: «Revolucionario será aquel que pueda revolucionarse a sí mismo».



No es de extrañar que científicos como Turing despreciaran a Wittgenstein por ser «un mero filósofo» que trabajaba para desdenar todo aquello que su mente no alcanzaba a entender. Desde luego, no parece muy lógico que a medida que una persona se desplaza a mayor velocidad, su medida relativa del tiempo se ralentice. Igual de absurda le debería parecer la codificación digital necesaria para poder computar mediante ordenadores —o máquinas de Turing—, dado que Wittgenstein entendía que la codificación digital es completamente insuficiente para captar la complejidad y los matices de la realidad que nos envuelve. En otras palabras, ningún conjunto de reglas, por complejas que sean, puede explicar la amplitud de la vida cognitiva humana. Por tanto, la única forma de reproducir un humano es... siendo humano.

Aunque durante su segunda etapa, el propio Wittgenstein echó por tierra su primera fase productiva, tampoco parece que los nuevos conceptos que propuso casaran bien con la inteligencia artificial. En esa segunda etapa defendió que el significado de una palabra varía en función del contexto. Lo vemos, por ejemplo, con la palabra «miedo»:

Esta película da mucho miedo.

¡Miedo me das!

Nos lo pasamos de miedo en la fiesta de la semana pasada.

Como se puede observar, el significado de la palabra «miedo» va transmutando en cada una de las frases anteriores. En primer lugar, usamos «miedo» en su sentido más literal, como sinónimo de terror o, según la definición de la Real Academia Española, «angustia por un riesgo real o imaginario». En la segunda expresión no se usa, obviamente, en el sentido de angustia por la presencia de un peligro, sino como un sentimiento de inquietud mucho más ligero, como el que produce el comportamiento de un interlocutor que se dispone a realizar una acción un tanto excéntrica o atrevida. Y, finalmente, en la tercera oración, usamos esa misma palabra en un sentido totalmente contrario al literal, es decir, como sinónimo de «muy bien» o «excelente».

Todo esto nos demuestra que el lenguaje es mucho más que una mera sintaxis. Es también una semántica que viene definida por el uso social que se hace del lenguaje. Y es en este terreno donde, según Wittgenstein, los ordenadores no pueden entrar.

Por tanto ¿puede una máquina «entender» el significado de una palabra? Es más, ¿puede una máquina entender el lenguaje oral o escrito, donde existe tal complejidad que los significados de las palabras van transmutando en función del contexto? Como veremos, la ciencia y la tecnología actual han podido responder a ambas preguntas. Pero eso fue mucho tiempo después de aquellas disputas que mantuvieron dos intelectos de altura como fueron Turing y Wittgenstein.

El avance de la inteligencia artificial y las guerras

Los conflictos bélicos siempre han espoleado el avance de la tecnología. Si durante la Segunda Guerra Mundial la inteligencia artificial progresó gracias a la necesidad de descifrar los códigos secretos de los nazis, al llegar la Guerra Fría lo hizo debido a la competencia entre las dos potencias mundiales.

En el verano de 1956, dos años después de la trágica muerte de Alan Turing, tuvo lugar un trascendental encuentro en la Universidad de Dartmouth College, en New Hampshire, Estados Unidos. En esta reunión, conocida como la conferencia de Dartmouth y organizada por el prominente informático estadounidense John McCarthy (1927-2011), se dieron cita once eminencias, entre ellas los estadounidenses Claude Shannon (1916-2001), criptógrafo y considerado el padre de la teoría de la información, Marvin Minsky (1927-2016), cofundador del laboratorio de inteligencia artificial y decano del Instituto Tecnológico de Massachusetts (el célebre MIT), el psicólogo Frank Rosenblatt (1928-1971), pionero de las redes neuronales, Herbert Simon (1916-2001), premio Nobel de Economía, y Allen Newell (1927-1992), experto en informática y psicología cognitiva. Ellos fueron los que consolidaron la columna vertebral de la inteligencia artificial mientras esta se propagaba por otras comunidades científicas de Occidente.

La idea de McCarthy era reunir a una decena de sabios durante dos meses para discutir temas como la cibernética, la teoría de autómatas o el procesamiento de información compleja. De

hecho, fue en esta reunión en la que se acuñó el término «inteligencia artificial» y esta fue la declaración fundacional del encuentro:

Proponemos que durante el verano de 1956 tenga lugar en el Dartmouth College en Hanover, New Hampshire, un estudio que dure dos meses, para diez personas. El estudio es para proceder sobre la base de la conjetura de que cada aspecto del aprendizaje o cualquier otra característica de la inteligencia puede, en principio, ser descrito con tanta precisión que puede fabricarse una máquina para simularlo. Se intentará averiguar cómo fabricar máquinas que utilicen el lenguaje, formen abstracciones y conceptos, resuelvan las clases de problemas ahora reservados para los seres humanos, y mejoren por sí mismas. Creemos que puede llevarse a cabo un avance significativo en uno o más de estos problemas si un grupo de científicos cuidadosamente seleccionados trabajan en ello conjuntamente durante un verano.

Durante esos sesenta días se debatieron dos modelos alternativos de inteligencia artificial, dos enfoques que fueron llamados *top-down* y *bottom-up*. La visión *top-down*, liderada por Newell y Simon (que aparecen en la fotografía superior de la pág. 61), argumentaba que la cognición era un fenómeno de alto nivel que, en cierta manera, podía separarse del cerebro. Por tanto, el cerebro no es más que un instrumento que puede reemplazarse por otras tecnologías de silicio capaces de realizar el procesamiento simbólico, que es lo que realmente posibilita la inteligencia. Este grupo defendía que, para crear inteligencia artificial, se debía trabajar en la elaboración de sistemas artificiales capaces de manipular los símbolos con las mismas reglas que lo hace el cerebro humano. En otras palabras, el enfoque *top-down* opinaba que los científicos debían olvidarse de reproducir la estructura física del cerebro para centrarse en las reglas y sistemas que permiten la manipulación y combinación de símbolos o sistemas simbólicos.

Por otra parte, el grupo que apoyaba el enfoque *bottom-up*, encabezado por Rosenblatt, argumentó que la estructura real del cerebro es primordial para la emergencia de la función cognitiva, y que, si queremos desarrollar sistemas de inteligencia artificial, primero hay que crear un hardware que imite la biología cerebral. El enfoque *bottom-up* iba muy en línea con las propuestas de Turing del ya nombrado modelo conexionista. De hecho, los científicos defensores de este enfoque avanzaron en su visión de inteligencia artificial mediante el desarrollo de la neurona artificial.

Las máquinas serán capaces, en veinte años, de hacer cualquier tarea que un hombre pueda hacer.
HERBERT A. SIMON en 1965

DE LOS SISTEMAS SIMBÓLICOS AL ENFOQUE CONEXIONISTA

Ambos enfoques se disputaron la supremacía durante los años cincuenta y principios de los sesenta, una época durante la cual hubo una sana competición entre las dos escuelas gracias a los fondos públicos vertidos en plena carrera tecnológica entre Estados Unidos y la antigua Unión Soviética, en el contexto de la Guerra Fría. En ese periodo, Newell y Simon, los defensores del enfoque *top-down*, desarrollaron un programa informático llamado Logic Theorist, que era capaz de demostrar teoremas nada triviales de lógica matemática.

Logic Theorist usaba lo que se conoce como *sistemas simbólicos*, los cuales fueron inventados por los matemáticos para dar sentido a expresiones sin referirse a convenciones arbitrarias. Por ejemplo, podemos decir que «ser un hombre» implica «ser mortal» y esta sentencia se puede formalizar mediante la expresión matemática $A \rightarrow B$, donde el símbolo «A» equivale a «ser un hombre», el símbolo « \rightarrow » significa «implica» y «B» equivale a «ser mortal». «Ser un hombre implica ser mortal» es una expresión arbitraria que se formaliza mediante la expresión $A \rightarrow B$. Una vez los términos arbitrarios están formalizados, es mucho más sencillo manipularlos y hacer operaciones con ellos desde un punto de vista informático-matemático.

Con el propósito de simplificar las operaciones matemáticas, los sistemas simbólicos parten de axiomas para construir teoremas con ayuda de reglas de derivación. La ventaja de los sistemas simbólicos es que, al ser sistemas formales perfectamente definidos y sin ambigüedades, su programación informática es relativamente sencilla. Veamos un ejemplo:

Sócrates es un hombre.

Todos los hombres son mortales.

Por tanto, como Sócrates es un hombre, es mortal.

Si lo formalizamos matemáticamente, estas sentencias quedarían de la siguiente manera:

A: Sócrates

B: hombre(s)

$A \rightarrow B$

C: mortal(es)

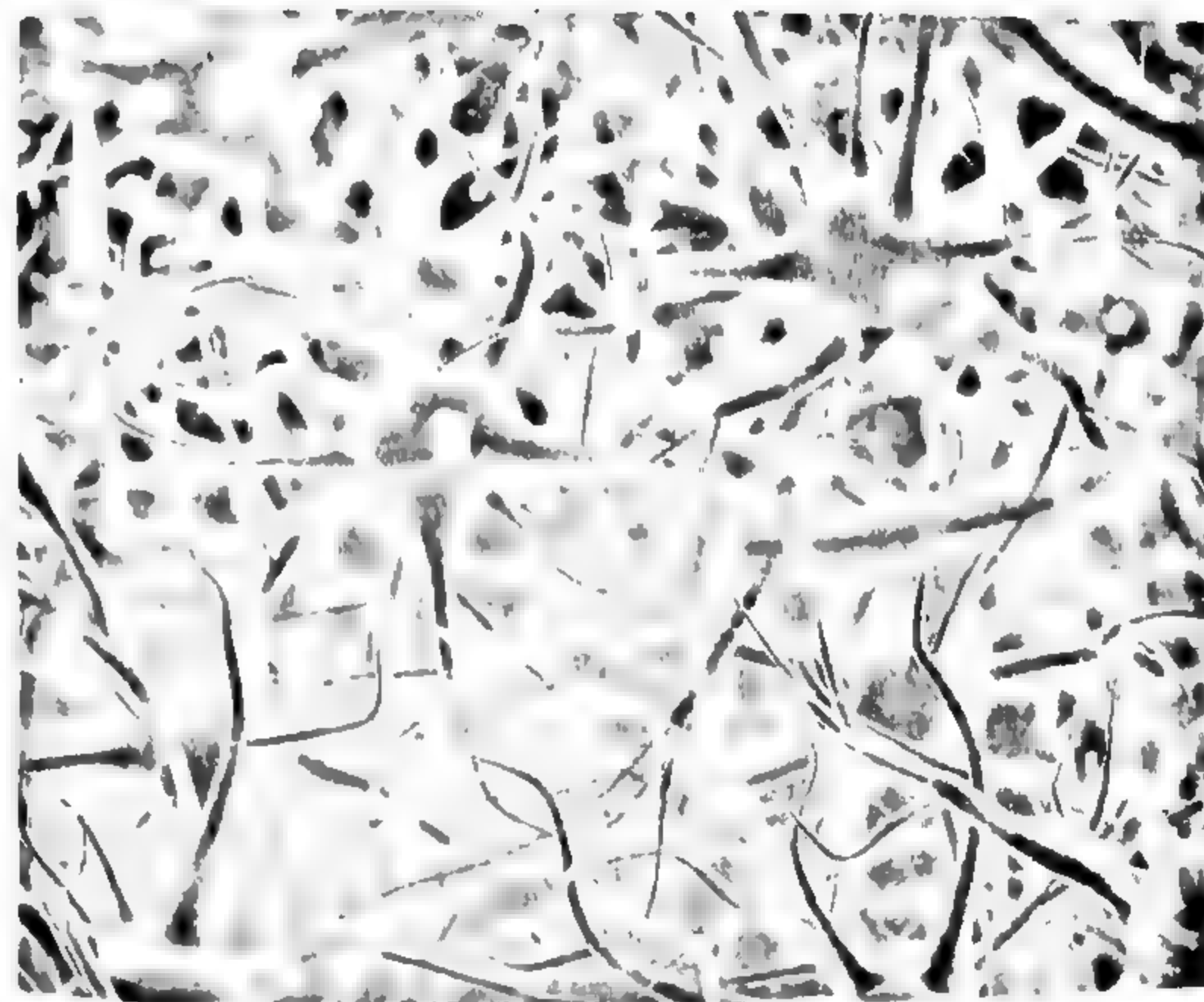
$B \rightarrow C$

Si $A \rightarrow B$ y $B \rightarrow C$, entonces $A \rightarrow C$, es decir, Sócrates es mortal.

En este caso, una regla de derivación conocida como «silogismo hipotético» nos permite concluir que $A \rightarrow C$, si es cierto que $A \rightarrow B$ y $B \rightarrow C$.

Sin embargo, la derivación automática y sistemática de teoremas a partir de los axiomas y de las reglas de derivación puede conducirnos a un número de combinaciones que, de nuevo, nos acerca peligrosamente al número de átomos del universo. Por ello, el programa Logic Theorist se valía de consideraciones heurísticas, es decir, un instrumento vagamente predictivo que ayuda a seleccionar las mejores derivaciones de entre todas las posibles, para identificar la secuencia correcta de derivaciones que deben realizarse sobre los axiomas, hasta llegar a la demostración de los teoremas.

Un ejemplo práctico sería el que se expone a continuación, en el que queremos saber si Sócrates es mortal o no lo es y conocemos los siguientes axiomas iniciales:



Herbert Simon (derecha) y Allen Newell en el despacho de Simon en el Carnegie-Mellon Institute en Pittsburgh (Pensilvania); la fotografía fue tomada en mayo de 1986. A la izquierda, recreación artística de una red neuronal.

A: Sócrates
 B: hinchas del Olympiacos
 C: griego
 D: hombre
 E: mortal
 $A \rightarrow C$
 $C \rightarrow D$
 $A \rightarrow D$
 $C \rightarrow B$
 $D \rightarrow E$

Y queremos saber si $A \rightarrow E$ es verdadero o falso mediante «fuerza» bruta, o lo que es lo mismo, probando todas las combinaciones posibles podemos tener:

$A \rightarrow C \rightarrow D \rightarrow E$
 $A \rightarrow C \rightarrow B$
 $A \rightarrow D \rightarrow E$

Es decir, hemos realizado siete operaciones matemáticas partiendo solo de cinco axiomas y usando únicamente una regla de derivación, el silogismo hipotético. Es fácil imaginarse que en escenarios más complejos, con más axiomas y más reglas de derivación en uso, el número de combinaciones posibles puede ser tan elevado que se tardarían años en obtener demostraciones concluyentes. Para atajar este problema, tal y como propusieron Simon y Newell, una consideración heurística (o como dicen los especialistas, «una heurística») nos habría avisado, en el ejemplo anterior, de que no vamos por el buen camino si para demostrar que alguien es mortal tenemos que empezar hablando de fútbol ($A \rightarrow C \rightarrow B$).

En la actualidad, más allá de los sistemas automáticos de demostración de teoremas matemáticos, tanto los sistemas simbólicos como las heurísticas son ampliamente utilizados en la resolución de problemas prácticos. Para ilustrar otro caso práctico de uso de heurísticas pondremos como ejemplo el juego del ajedrez. En un turno de ajedrez existen de media 37 posibles

movimientos. Un programa de ordenador que pretenda analizar una jugada con ocho turnos de profundidad deberá analizar 37^8 escenarios posibles, es decir 3512479453921 jugadas, o en otras palabras, más de 3,5 billones de jugadas. Si por cada jugada el ordenador dedicara un microsegundo a estudiarla, para analizar ocho niveles del juego (lo que para un jugador experto es bastante sencillo) un ordenador potente estaría «pensando»... ¡más de dos años y medio por turno! Como eso no es nada operativo, se requiere algún tipo de mejora en el método de análisis que permita acelerar el proceso, y ese avance lo constituyen las heurísticas. Su punto fuerte es que son reglas predictivas que ayudan al algoritmo a descartar aquellas jugadas que, por alguna razón, se percibe que van a conducir a situaciones muy desfavorables y que, por tanto, no hace falta continuar explorando. Solo con que las heurísticas eliminen el análisis de unas pocas jugadas absurdas, el ahorro en número de jugadas a analizar puede ser enorme. Muy basadas en la intuición del programador, las heurísticas son herramientas predictivas fundamentales en la mayoría de sistemas inteligentes y condicionan en gran medida la calidad de los mismos.

Un hardware que imite las conexiones neuronales: el enfoque conexionista

Defendido en Dartmouth por Rosenblatt, el enfoque conexionista es la alternativa al planteamiento de los sistemas simbólicos. Sus abanderados opinaban que, antes de plantearse el desarrollo de sistemas cognitivos, debe desarrollarse un hardware que compute de una forma similar a la que hace el cerebro, imitando las neuronas y las conexiones entre estas. Es decir, creando la neurona artificial y la red neuronal, respectivamente.

Ese concepto de red neuronal y neurona artificial ha pasado por diferentes ciclos de euforia-decepción a lo largo de su corta vida. Sus inicios se remontan al algoritmo conocido como *Threshold Logic Unit* (Unidad Lógica Umbral), propuesto por Warren McCulloch y Walter Pitts en la década de 1940 y que tuvo un

éxito abrumador. Una neurona artificial es, de hecho, la encapsulación de dicho algoritmo, que se define a continuación:

Entrada₁ → X_1

Entrada₂ → X_2

...

Entrada_i → X_i

si $\sum (X_i * \text{Peso}_i) > \text{Umbral}$;

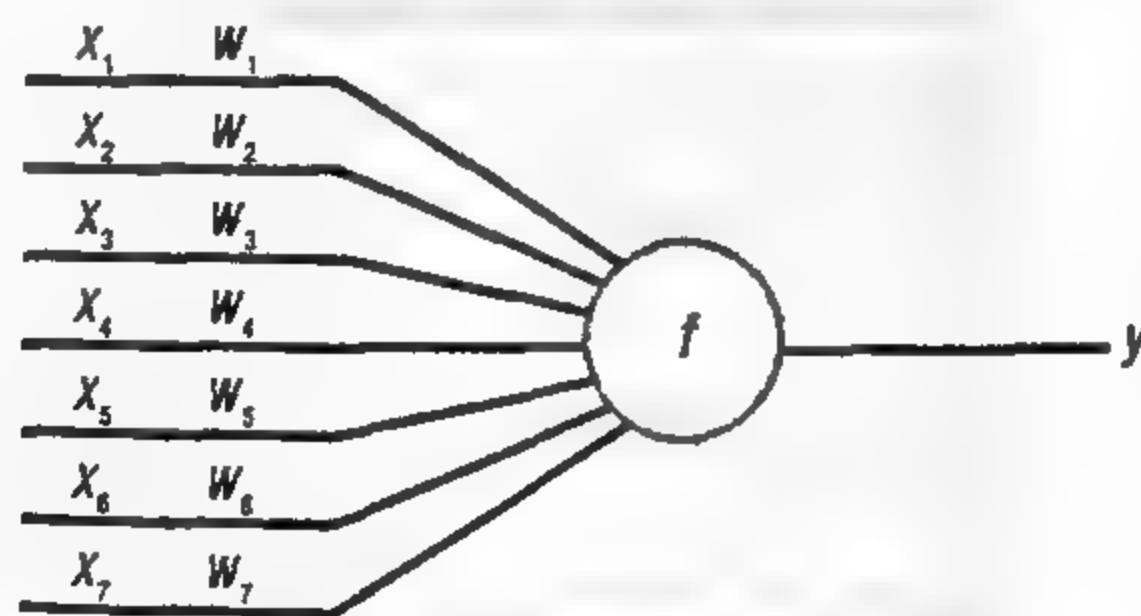
entonces Salida ← 1

si no Salida ← 0

Es decir, si computamos la suma ponderada de todas las entradas y el resultado excede un umbral, se propaga una señal, y si eso no sucede, no se propaga (figura 1).

Como se puede ver, una neurona es un elemento extremadamente simple: solo es capaz de abordar unas pocas operaciones aritméticas y una comparación. Este hecho facilitó la implementación de las neuronas artificiales en microchips y, acto seguido, la de redes neuronales completas en hardware a partir de finales de la década de 1990. Estas neuronas artificiales se usan en la actualidad para construir aparatos electrónicos de predicción,

FIG. 1



Esquema del algoritmo de una neurona artificial, donde X_i es la señal de entrada, W_i el peso, f es el algoritmo, e y , la señal de salida.

como por ejemplo un aparato que es capaz de detectar la causa del malestar de un bebé que llora.

El funcionamiento de una neurona artificial es parecido al de las neuronas naturales y, como se ha podido ver, es bastante sencillo. La dificultad de las redes neuronales radica principalmente en dos elementos que se deben ajustar y de los cuales dependerá que la red pueda realizar predicciones más o menos correctas. Se trata de los pesos de las diversas entradas y el umbral. El arduo proceso de ajustar estos valores para que, dada una serie de entradas, la neurona produzca la salida deseada, es lo que se conoce como «proceso de entrenamiento», o en términos psicológicos, el *aprendizaje*. La innovación del aprendizaje neuronal fue introducida por Frank Rosenblatt a finales de los años cincuenta. El invento de una neurona artificial capaz de ajustar los pesos y el umbral fue conocido como *perceptrón*.

Desde el punto de vista de la biología, el comportamiento de una neurona natural es prácticamente el mismo. Cada neurona tiene un conjunto de entradas por donde percibe señales eléctricas, provenientes de otras neuronas. Esto es lo que se conoce como *conexiones sinápticas*. A partir de las entradas, la neurona natural evalúa si estos estímulos superan un umbral de sensibilidad, siempre teniendo en cuenta que hay conexiones sinápticas que son más importantes que otras —los pesos de los que hablábamos antes—. Si se supera el umbral de sensibilidad, se propaga una señal eléctrica a través del axón o, lo que sería su equivalente en una neurona artificial, la salida.

Mediante esta relativa simplicidad, el perceptrón se planteó como una útil herramienta predictiva: dada una muestra, esa red neuronal artificial podía predecir si era de una clase (0) u otra (1). Un ejemplo clásico es el conocido como «Iris flor conjunto de datos», un conjunto de datos multivariante que hace referencia a muestras de tres especies botánicas: *Iris setosa*, *Iris versicolor* e *Iris virginica* (veánse las fotografías de la pág. 67). De cada una de las muestras recogidas se conocen cuatro parámetros: la longitud de los pétalos y su amplitud, y la longitud de los sépalos y su amplitud. El objetivo es que, dada una nueva muestra, la herramienta nos diga a qué especie botánica pertenece. En este

Las ciencias de la computación son una disciplina empírica. Cada nuevo computador que se construye es un experimento.

ALLEN NEWELL

caso, usaremos tres perceptrones, cada uno de ellos especializado en la detección de una de las tres especies botánicas. De tal manera que, si una nueva muestra es de la especie *Iris setosa*, solo uno de los perceptrones debería retornar 1, y los otros dos, 0.

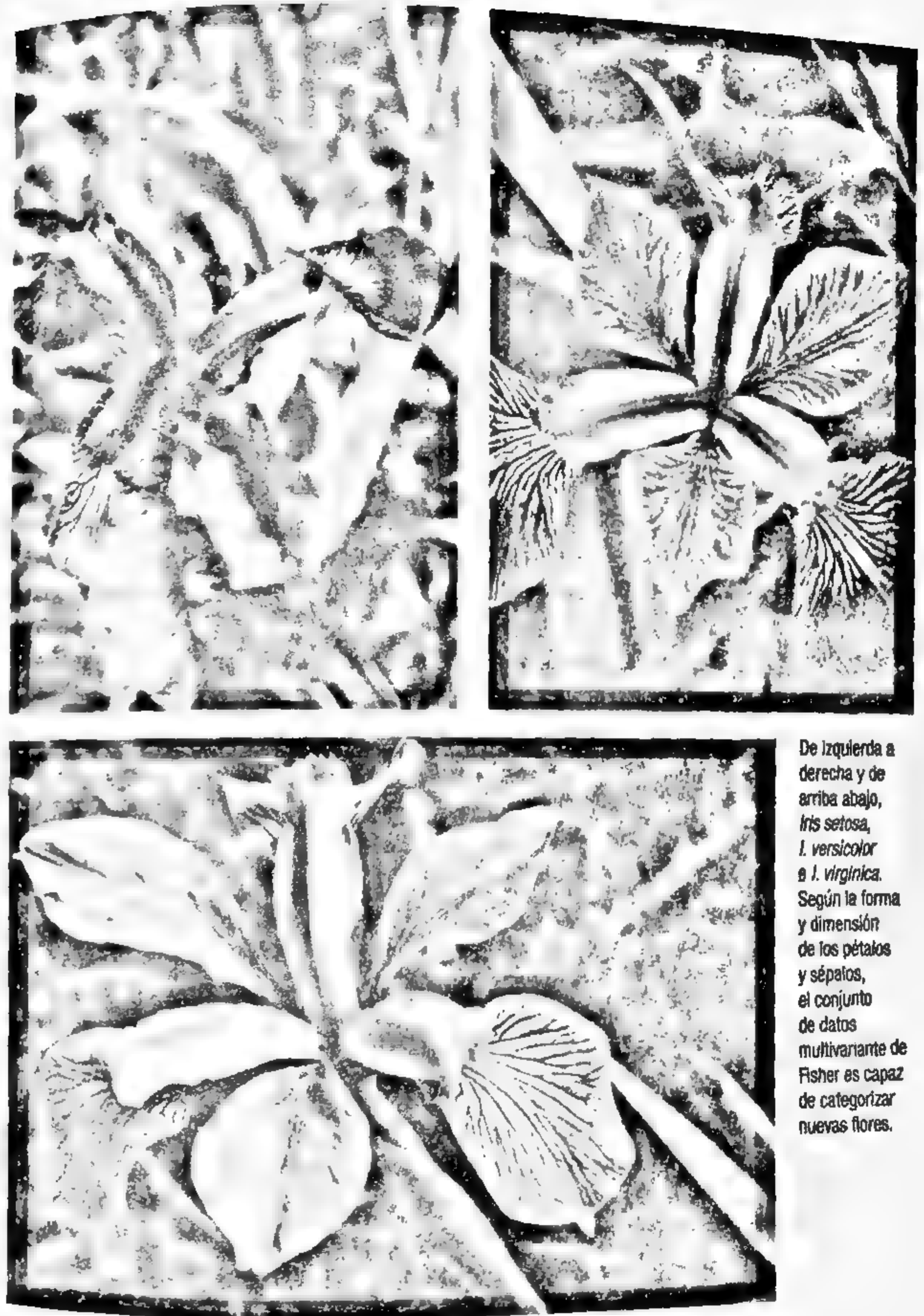
¿Podrían usarse herramientas del tipo estadístico para resolver el problema de las especies de *Iris*? En efecto, este problema es tan simple que también se puede resolver mediante herramientas estadísticas clásicas, como el análisis de componentes principales. Sin embargo, tanto las herramientas estadísticas como el perceptrón nos aportan dos maneras de razonar muy distintas y la más similar al razonamiento natural quizá sea la de este último.

En el caso de las herramientas estadísticas, hubiéramos obtenido reglas del tipo: «si los pétalos están entre tal y tal longitud y su anchura entre esta y esta otra, es probable que la muestra sea de la especie X». Mientras que la manera de razonar de un perceptrón es la siguiente: «si los pétalos están entre tal y tal longitud y tal y tal anchura, es probable que la muestra sea de la especie X, a no ser que tenga unos sépalos tan cortos que entonces los pétalos ya no importen, con lo que se podrá afirmar que es de la especie Y».

Es decir, mediante el sistema de ponderar el valor de las entradas para la toma de decisiones, hay información que puede pesar mucho menos que otra, pero, llegando a un valor extremo, esta entrada antes poco importante se convierte en muy relevante para la toma de la decisión.

¡PROBLEMAS A LA VISTA! SE PIERDE LA CONEXIÓN

La carrera intelectual entre los dos modelos se prolongó en el tiempo hasta que un acontecimiento imprevisto dio la ventaja a los del enfoque *top-down*. En 1969, Marvin Minsky y el sudafricano Seymour Papert, inventor del lenguaje de programación



De izquierda a derecha y de arriba abajo, *Iris setosa*, *I. versicolor* e *I. virginica*. Según la forma y dimensión de los pétalos y sépalos, el conjunto de datos multivariante de Fisher es capaz de categorizar nuevas flores.

Logo, publicaron el famoso libro *Perceptrons*, en el que se expone un simple problema de lógica booleana que una neurona artificial es incapaz de resolver, mientras que para la mente humana es algo trivial. Este hecho desencadenó una gran controversia, ya que se argumentó, e incomprensiblemente se convenció a la comunidad académica, que el enfoque conexionista no iba por el buen camino. ¿Cómo podía ser que no fuera capaz de resolver un problema tan concreto y trivial?

Veamos la cuestión detalladamente. El problema se generaliza enmarcado en la no-separabilidad lineal, pero Minsky y Papert lo ejemplificaron con el problema lógico del XOR o disyunción exclusiva. Si consideramos una situación donde las muestras puedan ser de dos categorías y cada una de ellas esté descrita por dos descriptores (y, por tanto, dos entradas), podríamos dibujar en el gráfico ocho muestras recogidas, como se aprecia en la figura 2.

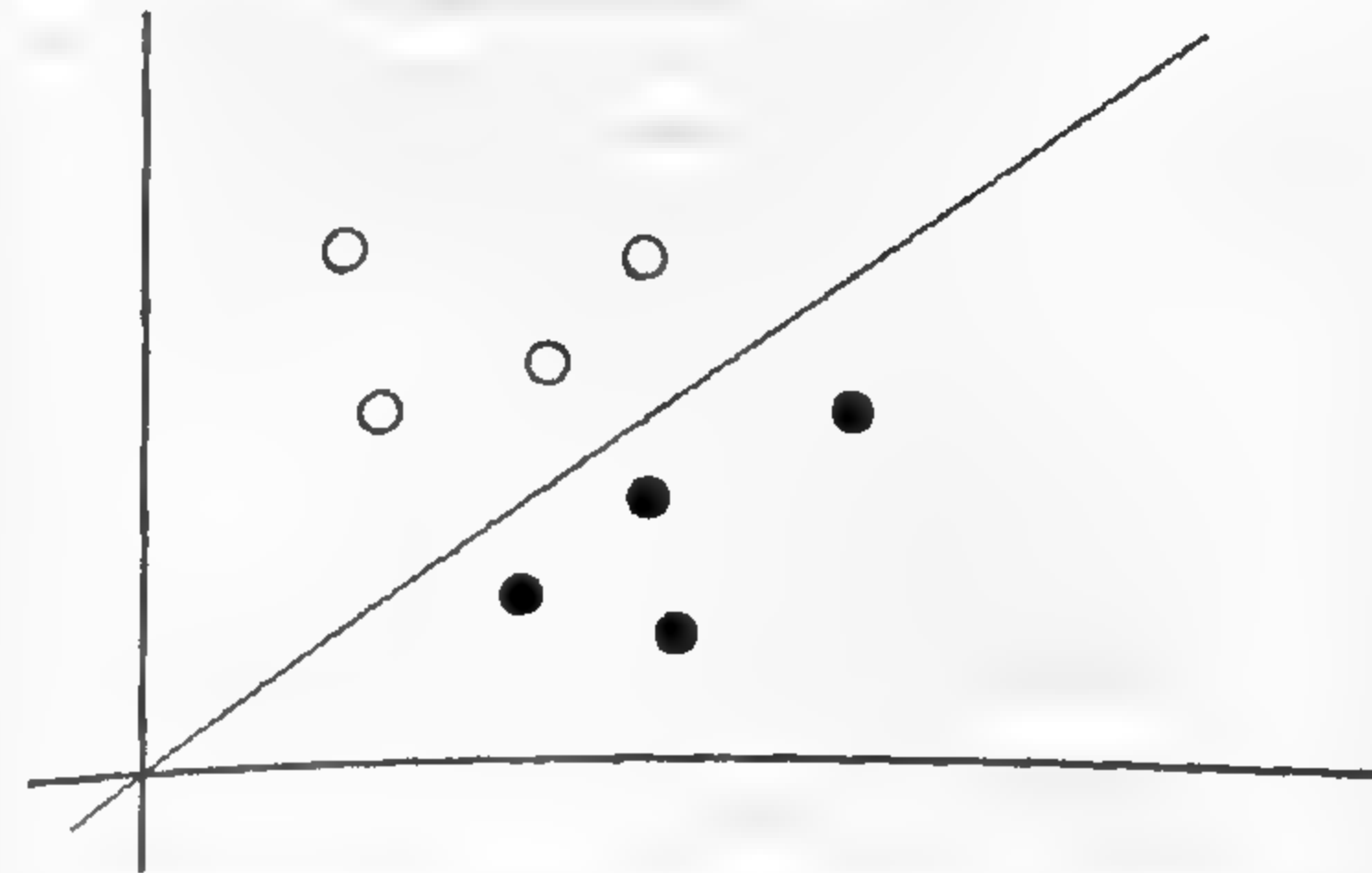
Tal como se muestra, es posible trazar una línea que separe ambas clases. Esto es justamente lo que hace un perceptrón cuando se ajustan el umbral y los pesos de cada entrada. Sin embargo, ¿qué pasa si analizamos el problema booleano de la XOR? La XOR es una operación lógica que cumple con la siguiente relación expuesta en la tabla:

Entrada A	Entrada B	Salida
0	0	0
0	1	1
1	0	1
1	1	0

Representada en dos dimensiones, esta función booleana quedaría como muestra la figura 3.

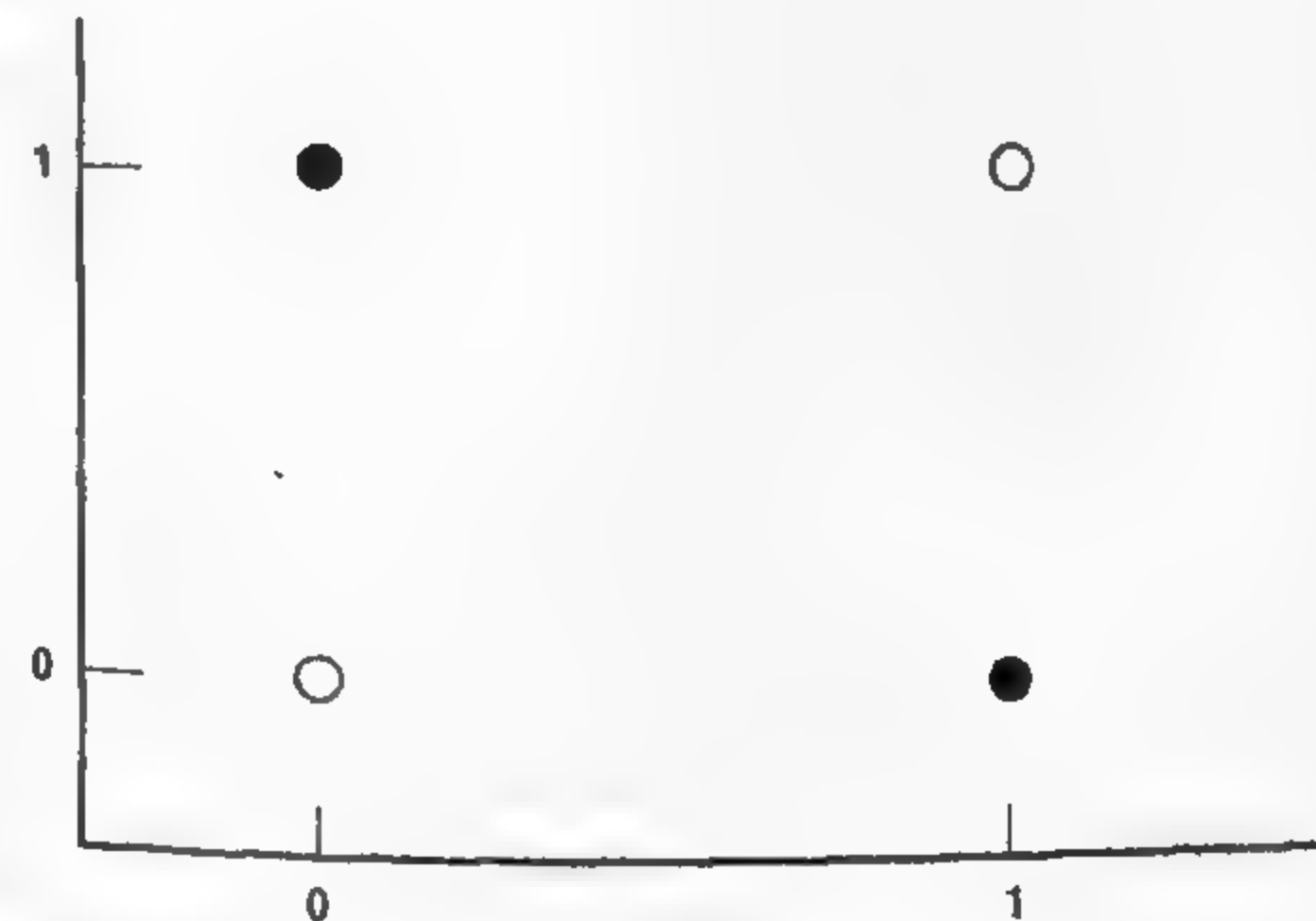
Como se puede ver, no existe ninguna recta que pueda separar los círculos blancos de los negros y, por tanto, estamos ante un problema no-separable linealmente. En definitiva, eso significa que un perceptrón no puede ser entrenado correctamente en la solución de un problema lógico tan simple como este. Este «im-

FIG. 2



Los círculos blancos representan las muestras de la categoría A, y los de color, las de la categoría B.

FIG. 3



Representación de la función booleana. Los círculos de color representan un 1, y los blancos, un 0.

previsto» provocó que los fondos destinados a la investigación de modelos *bottom-up* se redujeran de forma dramática. Por ello, la investigación en los años setenta se concentró en los sistemas simbólicos, a pesar de no estar exentos de grandes obstáculos. De hecho, el gran inconveniente al que los investigadores tuvieron que enfrentarse —y que solo ha sido resuelto recientemente— es el problema de la «inteligencia de fondo». Según argumentaban los científicos, desde la infancia cada uno de nosotros va atesorando un gran volumen de información básica acerca de cómo es el mundo. Por ejemplo, si decimos que una persona es un «cara dura» sabemos que difícilmente estamos hablando de una propiedad física de ese individuo, sino que es una expresión metafórica referida al hecho de ser un desconsiderado hacia los demás. Lo mismo sucede cuando en un día gélido decimos que hace un frío «polar» o que está «diluviando». Obviamente todo esto son metáforas, exageraciones o frases hechas que poco tienen que ver con su sentido literal. Sin embargo, cuando en una conversación entre humanos se utilizan estas expresiones —y lo cierto es que las utilizamos más de lo que nos damos cuenta y si no, ¡que me «cuelguen»!— las entendemos perfectamente debido a nuestra «inteligencia de fondo». Sin embargo, una máquina no tiene este catálogo de conocimientos del mundo que nos envuelve, y si recibe como entrada que hace un frío polar, lo más seguro es que establezca que la temperatura ambiente es de varios grados bajo cero, lo cual, seguramente, no será así.

En los años setenta y principios de los ochenta, el factor de la «inteligencia de fondo» fue fundamental para el desarrollo de la inteligencia artificial, ya que, en plena Guerra Fría, el interés principal de los estados occidentales —fundamentalmente, Estados Unidos— era generar sistemas inteligentes de traducción automática de documentos técnicos del ruso al inglés. Ante la imposibilidad de resolver el problema de la «inteligencia de fondo», las traducciones automáticas que se conseguían eran lamentables. Esto, sumado al tema de la no-separabilidad lineal, que había impactado directamente en la línea de flotación de la aproximación *bottom-up*, abrió las puertas de lo que hoy conocemos como el «invierno de la inteligencia artificial» (*AI Winter*).

LA INTELIGENCIA ARTIFICIAL, CRIOGENIZADA

En los quince años comprendidos entre 1970 y, aproximadamente, 1985, la disciplina de la inteligencia artificial se vio sumida en un largo invierno de producción científica, provocado por esos dos importantes obstáculos encontrados en las dos aproximaciones complementarias en las que se había estructurado la comunidad académica: la *top-down* y la *bottom-up*. Involuntariamente, los autores de ciencia ficción también contribuyeron a la decepción social generalizada al imaginar situaciones completamente irreales, como el caso ya comentado de la película 2001: *Una odisea del espacio*, que, como hoy sabemos, sería imposible que tuvieran lugar a corto y medio plazo. De hecho, cuando quedaba poco para el año 2001, ya quedó claro que una máquina como HAL 9000 era una quimera. Pero no solo los autores de ciencia ficción dejaron volar la imaginación, sino que también los científicos contribuyeron a crear grandes expectativas. Por ejemplo, Herbert Simon afirmó en 1965 que en veinte años tendríamos máquinas que podrían realizar cualquier trabajo que un humano fuera capaz de acometer, y Marvin Minsky publicó en 1970 en la revista *Life*, un medio generalista, que «dentro de entre tres y ocho años tendremos una máquina con una inteligencia similar a la de un ser humano medio». Lo cierto es que todo esto provocó un notable descenso de las subvenciones públicas destinadas a este campo de la ciencia y la tecnología, y a partir de aquí surgieron otros grandes problemas que entonces parecieron insuperables.

El primero de ellos fue el tema de la capacidad de cálculo y de los computadores en general. Un ejemplo: en aquella época se presentaron propuestas muy interesantes en el ámbito del procesamiento del lenguaje natural, pero estas solo podían manejar un vocabulario de veinte palabras, ya que en la memoria de los ordenadores de aquel tiempo no cabían más. Otro caso es el tema de la visión por computador. A final de los años setenta se estimó que, para detectar bordes y movimiento de igual forma que lo hace la retina humana, se necesitan unas mil millones de operaciones por segundo. Es decir 10^9 o, lo que es lo mismo en términos informáticos, 1 000 MIPS (millones de instruccio-

nes por segundo). Sin embargo, el ordenador más potente de la época y, por ende, más caro —costaba entre 5 y 8 millones de dólares— solo tenía una capacidad de 150 MIPS, el Cray-1. Hoy en día, para ponerlo en contexto, un teléfono inteligente de gama media tiene entre 15 000 y 20 000 MIPS, es decir, ¡centenares de miles de veces más potente que el ordenador que se usó para llevar al hombre a la Luna!

Otro de los dilemas que se puso de manifiesto en este invierno tecnológico fue la *explosión combinatoria*, una expresión matemática que constata que la mayoría de problemas a los que nuestra mente se enfrenta día a día o minuto a minuto, tienen no ya millones, sino, a menudo, billones de billones de combinaciones donde puede ubicarse la solución al problema. Por ejemplo, el número de posiciones legales en una partida de ajedrez se estima entre 10^{43} y 10^{60} . ¿Cómo va a un ordenador a explorar toda esta combinatoria para decidir la jugada óptima? Es más, ¿cómo funciona la mente humana, que puede resolver este problema sin necesidad explícita de analizar todas las jugadas? La cocina también muestra esa explosión combinatoria. Podemos inventarnos una nueva receta de cocina o incluso improvisar, y conseguir que salga un plato razonablemente apetecible. Sin embargo, si cuantificamos el conjunto de ingredientes que contiene una cocina doméstica, tenemos en cuenta los diferentes tiempos de cocción a los que sometemos cada ingrediente, así como el orden en que los combinamos y las cantidades que usamos de cada ingrediente, seguro que nos salen, de nuevo, billones de combinaciones posibles. Por tanto ¿cómo va a poder un ordenador improvisar una receta y que salga algo razonablemente apetecible?

El tercero de los problemas fue la versión generalizada de la ya comentada inteligencia de fondo. Cuando esta se puso de manifiesto, se planteó en un contexto de interpretación del lenguaje natural. Sin embargo, el problema se puede generalizar a cualquier otro ámbito. Si ahora mismo miro al cielo y veo algo blanco en forma de dragón, sé sin duda que no es un dragón, sino, probablemente, una nube que, por casualidad, ha adoptado esta forma de manera pasajera. Pero ¿cómo puede un ordenador distinguir entre la realidad y las apariencias? Los humanos lo ha-

remos por sentido común, pero los ordenadores no parece que puedan tener —o al menos, así se pensaba en los años setenta— esa capacidad natural que nos hace tan humanos.

El cuarto asunto, conocido como la *paradoja de Moravec*, es el problema de la percepción. El experto austriaco en robótica Hans Moravec (n. 1948), así como Minsky y otros científicos, plantearon que quizá el raciocinio humano necesita menos capacidad de computación que las habilidades que nosotros y muchos animales realizamos de forma inconsciente, como por ejemplo las sensoriales y motoras. Afirmaban que era asumible construir un sistema de inteligencia artificial con unas capacidades analíticas similares a las de un humano, por ejemplo, en el caso del ajedrez. Pero ¿qué pasa si veo entrar ahora mismo a mi esposa en la sala donde estoy escribiendo estas líneas?, ¿cómo voy a ser capaz de reconocerla en cuestión de milisegundos? Mi esposa es una persona entre 7 000 millones. Un ordenador tardaría mucho en contrastar todas las caras, mientras que un humano lo hace en pocos milisegundos. ¿Cómo es posible? Lo mismo pasa, decía Moravec, en otros ámbitos: hasta hace pocos años todavía no existían sistemas inteligentes capaces de hacer que las aeronaves aterricen de forma automática sin intervención humana. La complejidad técnica que esto conlleva, como es fácil de imaginar, es enorme. Sin embargo, un simple mosquito puede aterrizar y despegar en cualquier condición atmosférica con una fiabilidad del 100%. ¿Es más inteligente un mosquito que los sistemas de inteligencia artificial que incorporan las aeronaves modernas?

Por último, no podemos olvidar importantes obstáculos en la lógica formal disponible en el contexto histórico. Recordemos que la lógica booleana es una lógica de verdaderos y falsos, pero si yo me pregunto si soy una persona alta, ¿tengo que responder verdadero o falso? Los humanos no trabajamos en un esquema lógico binario, sino que todo es relativo. Un hombre que mida 1,70 m no se considera una persona alta. Pero si en su casa todos son de poca estatura, seguro que pasa por ser el alto de la familia. Esto sucede en todos y cada uno de los aspectos del lenguaje y el razonamiento. Como se suele decir, no todo es blanco o negro.

A consecuencia de todo ello, se provocó una debacle en la financiación de la investigación en esta materia. La institución de la época que más dinero invertía en la inteligencia artificial era la estadounidense DARPA (*Defense Advanced Research Projects Agency*), la agencia de investigación militar del país, la misma que creó internet y muchos otros inventos que han conformado la cotidianeidad del siglo XXI. Ante la decepción generalizada que provocó la inteligencia artificial en los años setenta y principios de los ochenta, DARPA cerró el grifo de la financiación a los grupos de investigación en la materia. Algunos programas se cancelaron abruptamente, entre ellos uno de reconocimiento de voz llevado a cabo por la Universidad Carnegie Mellon, situada en Pittsburgh, Pensilvania.

SISTEMAS EXPERTOS VERSUS REDES NEURONALES: EL REGRESO DE LA IA

Los ciclos, en la historia, suelen repetirse, como también fue el caso del resurgimiento del interés por la inteligencia artificial. Tras el parón, vino de nuevo un *boom* cuando, poco a poco, los académicos de las aproximaciones *top-down* y *bottom-up* fueron recibiendo dinero para sus propósitos. Eso sí, cambiando los nombres para romper con la sensación de fracaso percibida por la sociedad unos años atrás. Los de la aproximación *top-down* se identificaron con el concepto de sistemas expertos, mientras que los de la aproximación *bottom-up* lo hicieron con las redes neuronales.

Un sistema experto es un programa de ordenador que trata de recoger el máximo conocimiento sobre un dominio muy concreto para tratar de emular el comportamiento de un humano avezado en la materia mediante el uso de un sistema lógico. Por ejemplo, el sistema experto MYCIN fue desarrollado a principios de la década de 1970 en la Universidad de Stanford con el fin de identificar las bacterias causantes de infecciones severas y poder recomendar el mejor tratamiento en cada caso. MYCIN trataba de aunar todo el conocimiento que tiene

un médico experto en infecciones bacterianas y, mediante un procedimiento lógico, combinaba las evidencias del paciente para sugerir el tratamiento óptimo. El sistema llegó a tener una tasa de aciertos del 69%, superior a la de la mayoría de médicos no especializados. Pero en la práctica no llegó a usarse, entre otras cosas por el dilema que representaba delegar en una máquina la decisión de un diagnóstico. ¿Quién cargaría con la responsabilidad en caso de error?

Pero, probablemente, el sistema experto más famoso fue el XCON, desarrollado por la Universidad Carnegie Mellon para la empresa DEC (*Digital Equipment Corporation*). XCON —acrónimo de eXpert CONfigurer— ayudaba a configurar los equipos informáticos adquiridos por los clientes de DEC en función de sus necesidades y sus componentes hardware. Según se ha documentado, este sistema experto ahorró 40 millones de dólares anuales entre 1980 y 1986, una cifra nada trivial para aquella época en la que el desarrollo tecnológico era tan incipiente.

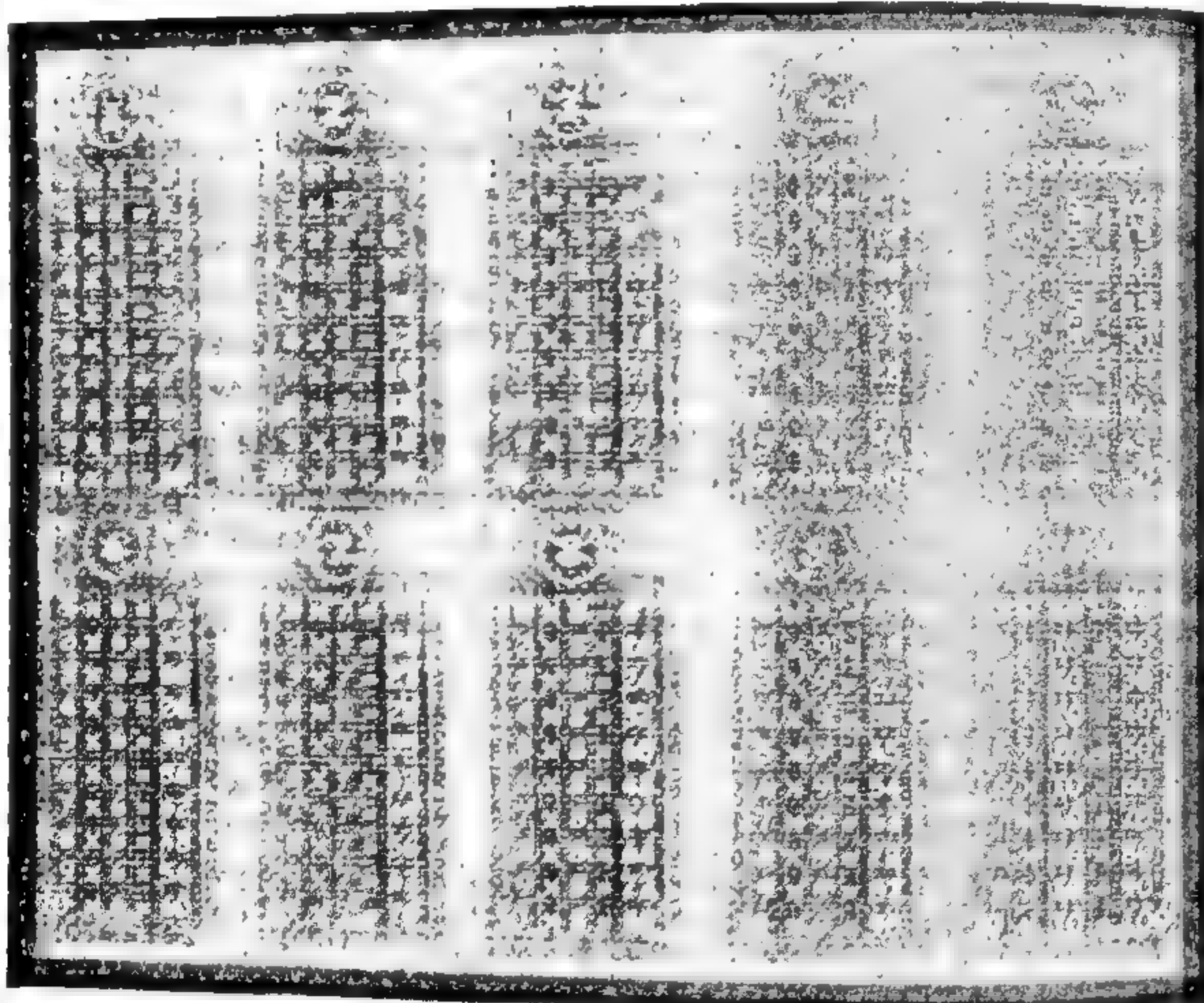
Por su parte, el enfoque conexionista también fue avanzando progresivamente y el problema de la no-separabilidad lineal fue resuelto mediante algo tan obvio y natural que *a posteriori* no se alcanzaba a comprender cómo los investigadores no habían dado con la solución con anterioridad. De hecho, la naturaleza ya había descubierto la respuesta millones de años atrás. La solución era conectar diversos perceptrones entre sí para formar lo que se conoce como redes neuronales (figura 4).

Una red neuronal puede ser tan compleja como se desee, con tantas capas ocultas como el programador considere oportuno, e incluso con conexiones que pueden ir de delante atrás para simular una especie de memoria. De hecho, se han llegado a construir redes neuronales con 300 000 neuronas, que es el mismo número de neuronas que contiene el sistema nervioso de una lombriz de tierra. La red neuronal de la figura 4 se denomina *feed-forward* (algo así como «alimentado hacia delante»), ya que el flujo de los datos siempre es de izquierda a derecha y no se forman ciclos entre las conexiones sinápticas.

En una red neuronal el aprendizaje se complica y por ello los ingenieros han ideado un gran número de métodos de aprendiza-

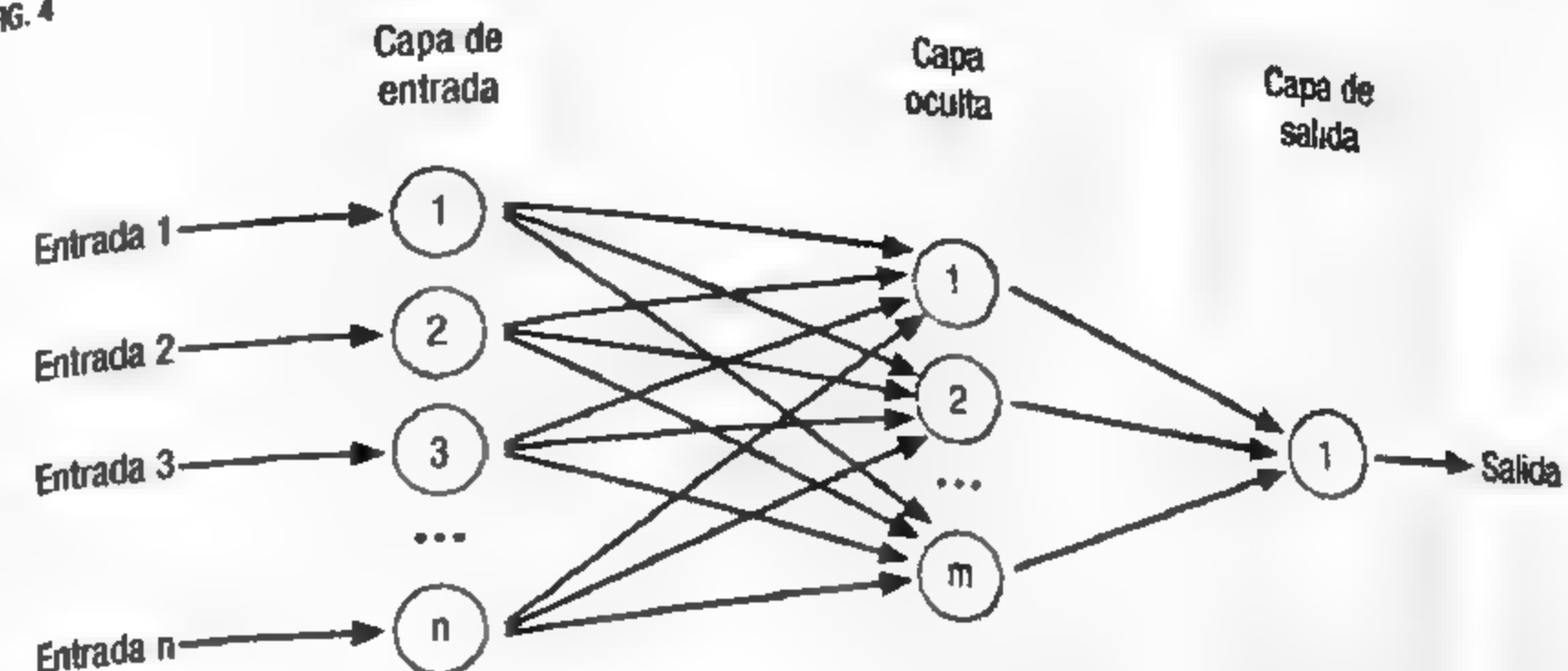
LA TRAMPA DEL SOBREENTRENAMIENTO

Un sistema de predicción basado en aprendizaje artificial deduce sus predicciones de la generalización que es capaz de hacer a partir de casos pasados. Por tanto, cuando el sistema no puede generalizar, pierde su utilidad. Cuando el proceso de entrenamiento se repite demasiadas veces, llega un punto en que el ajuste es tan exacto y está tan adaptado a las muestras de entrenamiento que el sistema las ha memorizado y ya no basará sus predicciones en una generalización sino en la memorización. Cuando esto sucede, el sistema solo es capaz de realizar predicciones correctas cuando se le introducen muestras del conjunto de entrenamiento, mientras que al introducir una muestra distinta para que realice una predicción, esta será incorrecta. Entonces se dirá que el sistema está sobreentrenado. De alguna manera, sucedería como un niño que, en vez de aprender a multiplicar, memoriza solamente las tablas. Al preguntarle por una de las operaciones que ha memorizado, contestará correctamente sin dudarlo, pero no sucederá lo mismo si se le pregunta por una nueva multiplicación que no forma parte de las tablas.



Las tablas de multiplicar son un buen ejemplo de aprendizaje por memorización.

FIG. 4



Una red neuronal formada por tres capas neuronales. La primera es la de entrada, la segunda es la oculta y la última, la de salida.

je. Uno de los más simples es el método de la retropropagación (en inglés *back-propagation*), que además da nombre a las redes neuronales que usan este método.

El método de la retropropagación consiste en minimizar el error de salida de la red neuronal a base de ajustar de derecha a izquierda los pesos de entrada de las conexiones sinápticas de las neuronas siguiendo el método del descenso de gradiente. Es decir, primero se dan valores aleatorios a todos los pesos de todas las conexiones de la red. A continuación, se introduce una muestra con el valor a predecir conocido —y, por tanto, se dice que es una muestra de entrenamiento—. Como es de esperar, el resultado que retornan las neuronas de salida será un valor aleatorio. A partir de aquí, comenzando por las neuronas más cercanas a la salida y terminando en la entrada, se empiezan a ajustar los valores de los pesos de las conexiones, para que el valor de la neurona de salida se acerque al valor real conocido.

Este procedimiento se repite centenares o miles de veces con todas las muestras de entrenamiento. Cuando ya se ha hecho todo el proceso con todas las muestras, se dice que ha pasado una época. A continuación, se puede repetir el proceso otra

LA INTRODUCCIÓN DE NUEVAS DIMENSIONES EN EL PROBLEMA DEL XOR

Como no podía ser de otra manera, el problema de la no-separabilidad lineal fue elegantemente solucionado mediante la introducción de las redes neuronales. Ya comentamos que el problema fue ilustrado por Minsky y Papert con el problema del operador lógico simbolizado como XOR: como demuestra la figura 1, no existe ninguna recta que pueda separar las muestras «0» y «1» en su representación en el espacio. Veamos cómo se resuelve el problema de la no-separación lineal, ilustrándolo también con el problema de la XOR y tal como lo haría una red neuronal.

Un sistema multicapas

La esencia de una red neuronal es la introducción de múltiples capas de neuronas en el sistema. Estas capas aportan la capacidad analítica-algebraica de añadir más dimensiones a un problema. Por tanto, lo que se hace para resolver el XOR es añadir una tercera dimensión a la gráfica anterior que llamaremos «Z», y $Z = A \times B$, como muestra la figura 2. En este caso, gracias a la dimensión «Z», ya existe, no una recta porque estamos en un espacio de tres dimensiones, pero sí un plano, que puede separar los círculos blancos de los negros. En la figura 3, los dos círculos blancos estarían situados entre el lector y el plano, mientras que los dos de color estarían entre el plano y el fondo de la hoja. Por tanto, ahora podemos decir que gracias a este «truco», el problema del XOR ya es linealmente separable.

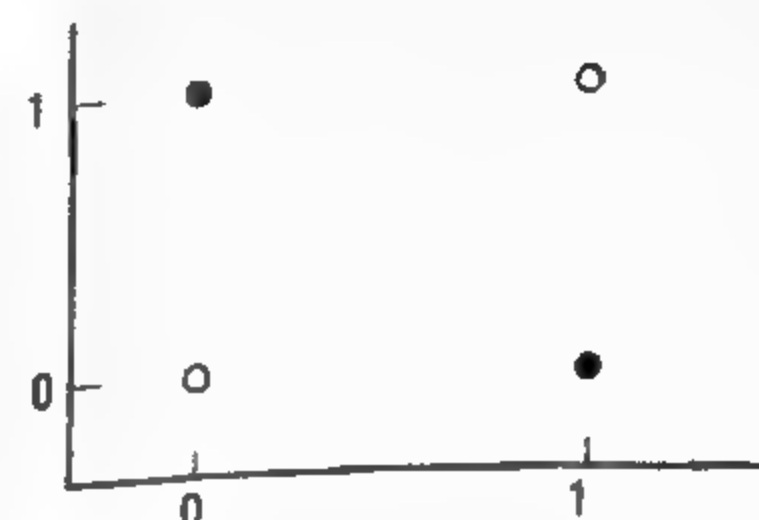
Un proceso cognitivo demasiado complejo

Lo que queda claro es que una red neuronal no es más que la concatenación y la combinación de múltiples perceptrones, lo cual aumenta de forma muy notable la capacidad cognitiva del invento. Esto, sin embargo, presenta un problema, y es que una vez que tenemos entrenada una red neuronal, su proceso cognitivo es incomprensible incluso para un experto en la materia. Un aspecto al que no se le dio importancia hasta que las redes neuronales fueron introduciéndose plenamente en la resolución de problemas reales. Por ejemplo, si se usa una red neuronal para controlar el sistema de frenado ABS de un coche, los ingenieros querrán asegurarse de que entienden hasta el último detalle del razonamiento de la red neuronal para poder garantizar que los frenos no van a fallar ante cualquiera de las miles de situaciones diferentes de frenado en las que pudieran verse implicados.

Algoritmos de aprendizaje supervisado

Por este importante detalle, diversos teóricos de la computación trabajaban intensivamente entre finales de la década de 1990 y principios del siglo XXI para diseñar nuevos métodos computacionales capaces de solventar o mitigar este efecto de las redes neuronales. La solución final vino de la mano de Vladimir Vapnik y su equipo, a principios del siglo XXI, desarrolladores en la famosa empresa de telecomunicaciones y equipos electrónicos AT&T Bell Labs. Vapnik ideó las máquinas de soporte vectorial (*Support Vector Machine*, SVM por sus siglas en inglés) que, principalmente, se basan en introducir nuevas dimensiones artificiales de antemano y de forma controlada a un problema no-separable linealmente para que ahora, gracias a las nuevas dimensiones artificiales introducidas, sí lo sea.

FIG. 1



Representación bidimensional de la función lógica XOR.

Añadimos un tercer eje (Z), que es fruto de multiplicar A x B.

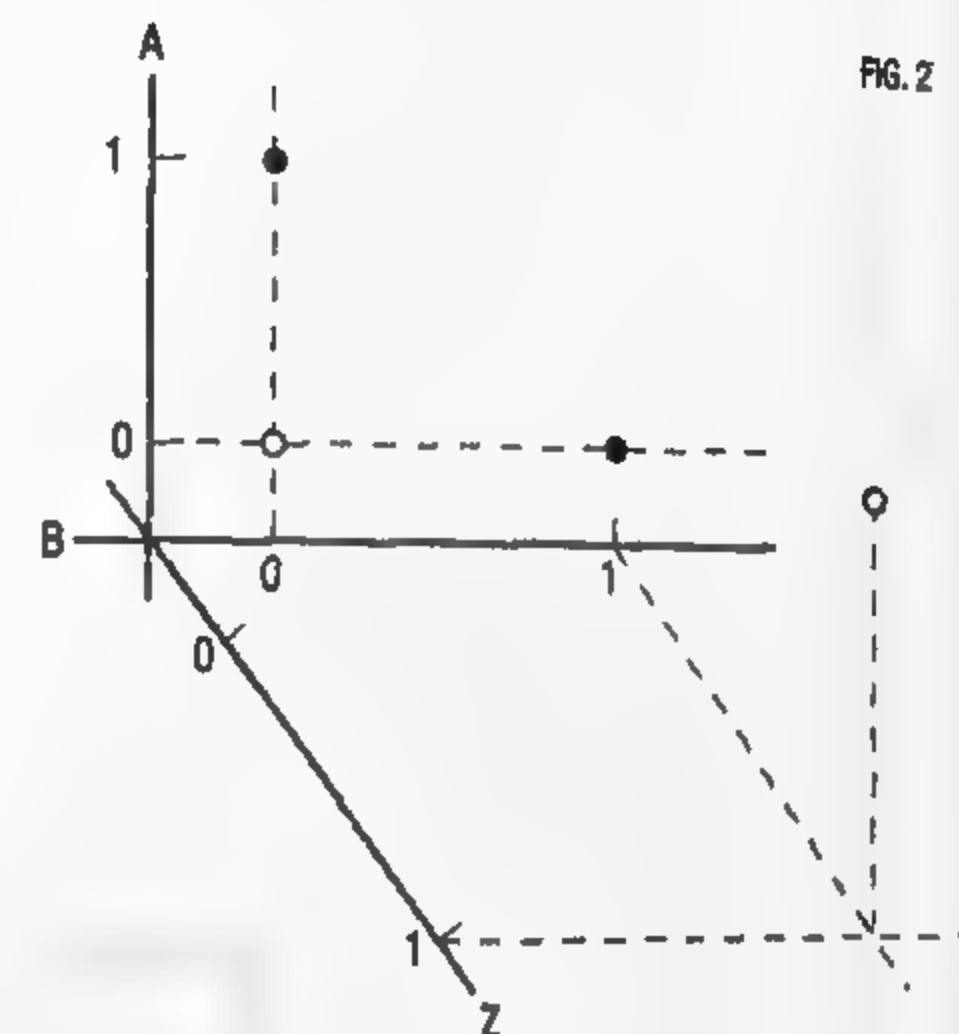
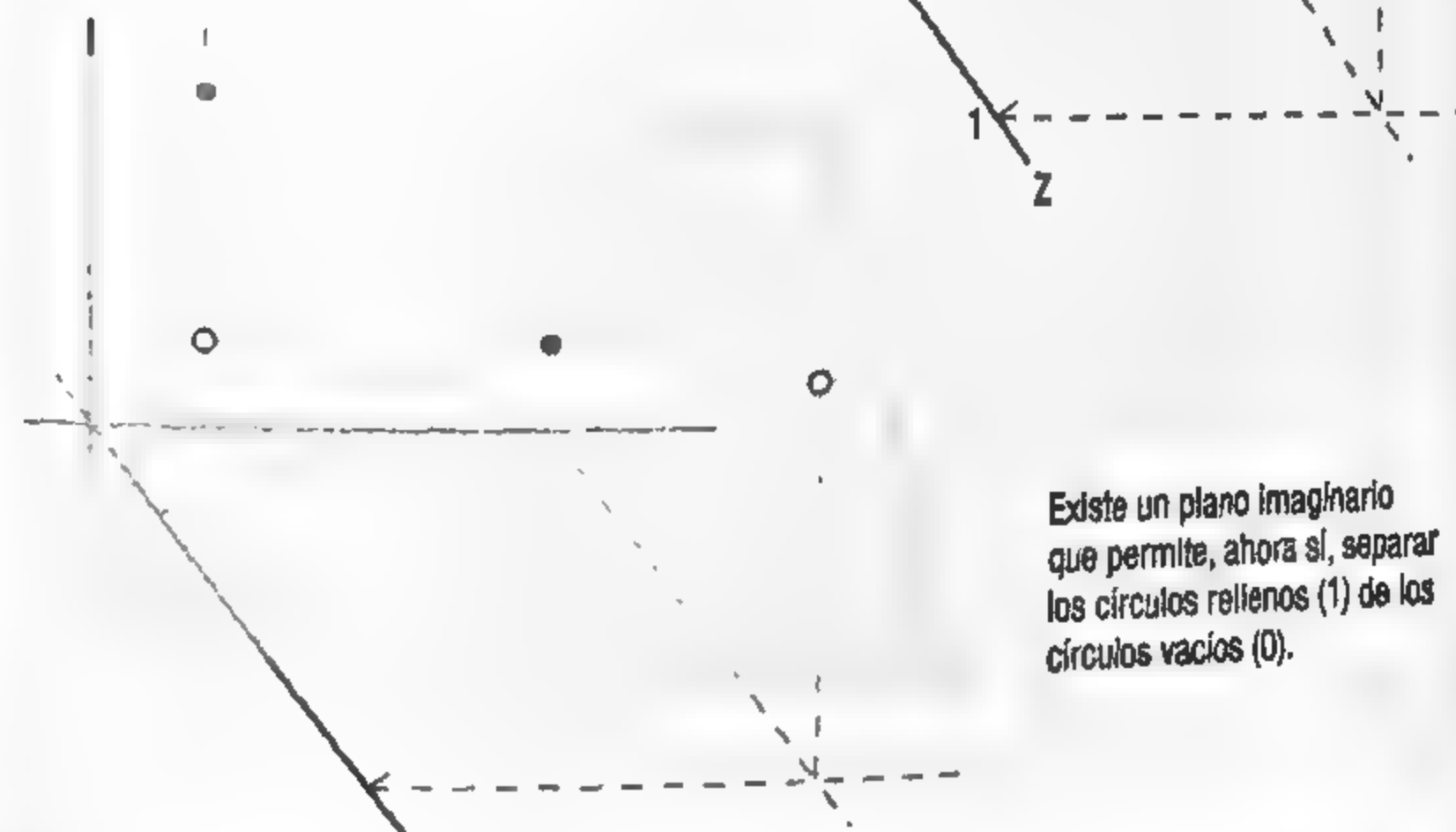


FIG. 2

FIG. 3

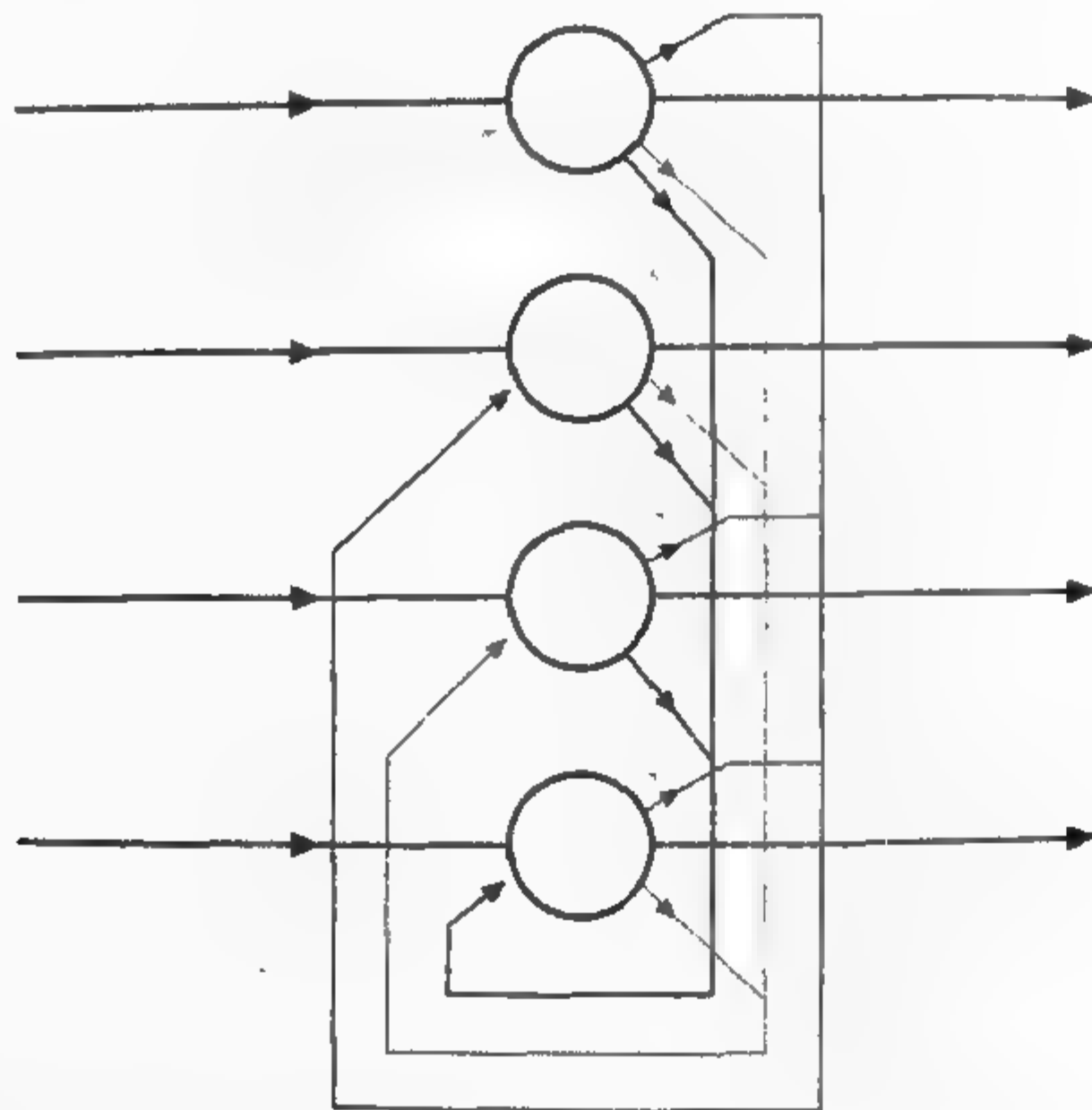


Existe un plano imaginario que permite, ahora sí, separar los círculos rellenos (1) de los círculos vacíos (0).

época entera, con las mismas muestras de entrenamiento. Un proceso de aprendizaje normal consta de unas pocas decenas de épocas. Nótese que este proceso es similar al psicológico de aprendizaje, en el que se aprende por repetidas observaciones de los mismos datos.

La arquitectura de una red neuronal y los métodos de aprendizaje fueron sofisticándose a lo largo de los años y surgieron un gran número de tipos de redes neuronales diferentes para aplicaciones de la vida real muy diversas. Una de las redes neuronales artificiales más usadas en la actualidad son las redes de Hopfield (por su inventor, el científico estadounidense John Hopfield, que las creó en 1982), que implementan un tipo de dispositivo de memoria llamado *memoria asociativa* (figura 5).

FIG. 5



Esquema de una red neuronal de Hopfield de cuatro nodos. Las unidades o nodos de una red neuronal Hopfield son binarias, de manera que tienen dos valores posibles, que quedarán determinados en función de si las unidades superan o no un valor concreto.

En una memoria asociativa, la información está ordenada según el contenido y, por tanto, es necesario indicarle el contenido al que queremos acceder, en vez de una posición física-electrónica, como se hace en un disco duro o una memoria RAM de un ordenador.

Otro tipo de red neuronal muy usado en la realidad son las redes o mapas de Kohonen (por el profesor finlandés Teuvo Kohonen) o autoorganizados. En este caso, la red neuronal incorpora una innovación: el aprendizaje no es supervisado sino que es la misma red la que va autoaprendiendo de sus errores a medida que va trabajando.

Con todo esto, la Guerra Fría llegó a su fin y, de nuevo, la inteligencia artificial volvió a sufrir otra era glacial, mucho menos intensa que la acaecida entre los años setenta y ochenta. De hecho, fue una suerte de calma antes de la tempestad, entendiendo esa tempestad como el avance final que ha provocado que hoy en día la inteligencia artificial, probablemente sin saberlo, esté gobernando prácticamente todos los aspectos de nuestra vida cotidiana.

La inteligencia artificial en el presente

Tras superar un sinfín de obstáculos, hoy en día la inteligencia artificial es parte indisoluble de nuestra vida cotidiana. Veamos cuál es su alcance en el momento actual y qué tecnologías han sido imprescindibles para llegar hasta este punto.

Como si de una burbuja económica se tratara, después de ese parón acaecido entre 1970 y 1985, y tras solo dos años de apogeo, la inteligencia artificial volvió a caer en un nuevo invierno en el periodo comprendido entre 1987 y 1993. Algunos de los motivos fueron de índole económica, pero también hubo problemas técnicos y dilemas filosóficos. Paradójicamente, la notable aceleración que experimentó la computación de propósito general en este periodo fue la causante de los problemas financieros que afectarían a la IA. En aquella época, IBM y Apple lanzaron los primeros ordenadores personales, de gran éxito comercial entre el público general. Pero ello supuso que un ordenador que estaba especialmente diseñado para soportar sistemas de inteligencia artificial, de nombre Lisp Machine, quedara desbankado de la noche a la mañana y, con él, todo un negocio de más de 500 millones de dólares.

Por otro lado, también hubo motivos técnicos, como los relacionados con los sistemas expertos que, recordemos, son esos programas de ordenador que tratan de recoger el máximo conocimiento sobre un dominio muy concreto para tratar de emular el comportamiento de un experto en una determinada materia

mediante el uso de un sistema lógico. El problema con los sistemas expertos fue que, con el tiempo, se demostró que eran muy caros de mantener y que, además, no podían aprender, con lo cual, rápidamente se convertían en obsoletos y, por tanto, inútiles. Para agravar el problema, a menudo esos sistemas solían dar respuestas grotescas ante situaciones anómalas, lo que les descreditaba y les restaba el prestigio ganado hasta ese momento. Por si fuera poco, a las dificultades económicas y técnicas se añadieron dilemas de tipo técnico, como es el caso del problema de la cualificación, que hace referencia a la imposibilidad manifiesta de poder listar de forma exhaustiva todas las precondiciones necesarias para poder ejecutar una tarea. Este hecho no tendría ninguna importancia si no fuera porque, tal y como estaban concebidos, los sistemas expertos necesitaban conocer de antemano todo el contexto para resolver un nuevo problema. Pongamos un ejemplo: imaginemos que queremos cruzar un río con una piragua. Si queremos describir todas las condiciones necesarias para acometer nuestro objetivo, podemos empezar aseverando que necesitamos un río, que, por supuesto, contenga agua, y una piragua capaz de flotar. Y si queremos remar, deberemos tener brazos, remos, la posibilidad de respirar, fuerzas suficientes para no desistir a medio camino, etc. En definitiva, las precondiciones pueden ser infinitas y los sistemas expertos las necesitan todas para hacer bien su trabajo. La conclusión obvia a todo esto es que, en lo referente a la configuración de ordenadores, fuera de entornos muy, pero que muy controlados, como es el caso ya comentado del XCON, los sistemas expertos no pueden ser programados para que añadan algún valor real.

La cuestión es que, tanto por los problemas del hardware con el Lisp Machine como por los del software con los sistemas expertos, la inteligencia artificial volvió a caer de nuevo en una profunda depresión. A todo esto se añadió otra complicación más, en este caso en el plano filosófico, conocida como la *Nouvelle AI*, que fue una suerte de recaída en la ya nombrada paradoja de Moravec. La *Nouvelle AI* —AI viene de las siglas del inglés de *Artificial Intelligence*— es una vuelta a la visión más antropocéntrica de la inteligencia artificial que tuvo lugar

LISP MACHINE, EL LENGUAJE DE PROGRAMACIÓN DE LA IA

Lisp es el nombre de un lenguaje de programación que se ha usado históricamente para programar sistemas de inteligencia artificial de forma mayoritaria. A finales de los años setenta, la compañía Lisp Machines, Inc. empezó a fabricar y distribuir un computador que ejecutaba instrucciones en Lisp de forma nativa (es decir, en base a un código nativo, configurado para un procesador específico). La máquina entendía ese lenguaje sin necesidad de trasladarlo previamente a ningún otro lenguaje de nivel más bajo, que es lo que sucede en la mayoría de los ordenadores que utilizamos hoy en día. Gracias a ello, el ordenador podía ejecutar de forma muy rápida instrucciones en Lisp y los ingenieros se podían comunicar directamente con el computador usando este lenguaje, sin necesidad de trasladar las instrucciones a otros entornos para hacer su ejecución más eficiente.

Los PC acaban con las Lisp Machines

Visto su éxito inicial, varias compañías fabricaron y vendieron Lisp Machines, entre ellos Symbolics (en la imagen, un modelo de esta compañía), Texas Instruments o Xerox. Sin embargo, muchas de ellas tuvieron que cerrar u olvidarse de continuar fabricando este computador durante la denominada «revolución de la microcomputación», liderada por IBM y Apple. Es decir, cuando empezó a generalizarse la posesión de ordenadores personales como los que hoy en día todos tenemos en casa. Y es que llegó un momento en el que con los PC se podían ejecutar programas en Lisp de forma más rápida que con la misma Lisp Machine, sin necesidad de ningún hardware específico, y a unos costes muchísimo menores, lo que, lógicamente, marcó su fin en el mercado.

Symbolics 3640 Lisp Machine.



en los años ochenta. Liderados por el ingeniero Rodney Brooks, del MIT (que aparece en la imagen de la página contigua junto a uno de sus robots humanoides), una serie de pensadores del campo de la cibernética y la teoría de control defendían acérrimamente que la inteligencia artificial solo puede emerger a partir de la percepción e interacción de la máquina con el mundo que nos rodea. En otras palabras, que es importante superar un test de Turing con un hardware que imite la biología humana. Por tanto, todos estos autores desechaban todos los avances en el campo simbólico realizados hasta el momento. En realidad, se sentían herederos de la aproximación más radical del *bottom-up*. Hecho que, unido al desprestigio en el que se vieron sumidos los sistemas simbólicos del momento, los sistemas expertos, acabó de agravar la crisis de la inteligencia artificial. Afortunadamente, y a pesar de la escasez de fondos para la investigación que estas disertaciones contribuyeron a ahondar, la rama de la Nouvelle AI dio unos no despreciables frutos en el campo de la robótica insectoide y humanoide.

EL HOMBRE CONTRA LA MÁQUINA: DEEP BLUE VERSUS GARRI KASPÁROV

Una vez superada esta segunda crisis, sin hacer demasiado ruido mediático —los investigadores ya habían aprendido la lección— la inteligencia artificial experimentó un desarrollo exponencial y a la vez espectacular. La disciplina fue avanzando en paralelo en sus múltiples subcampos: el razonamiento simbólico, los modelos conexionistas, la robótica y cibernética, etc. Hasta que llegó un momento clave que la devolvió al campo mediático y, por tanto, al público general: la llegada de Deep Blue. Por primera vez no solo se superaba con éxito un test de Turing, sino que una máquina vencía al mayor especialista mundial en un juego intelectual como es el ajedrez.

El ajedrez es un clásico problema combinatorio al que, desde los principios de la informática, se ha intentado aplicar técnicas inteligentes para vencer a jugadores humanos. Ya vimos en capí-



Rodney Brooks posando junto a un robot desarrollado por él en su despacho del MIT Computer Science and Artificial Intelligence Laboratory.

tulos anteriores intentos en ese sentido como los de científicos de la talla de Torres Quevedo e incluso fraudes como el del supuesto autómatas conocido como El Turco. ¿Por qué es tan difícil fingir la inteligencia en problemas como el del ajedrez? Imaginémonos que introducimos en un ordenador las reglas del ajedrez y este construye el conjunto de todas las jugadas posibles. Se-

La táctica implica cálculos que comprometen al cerebro humano, pero cuando los descompones, resultan ser la parte más simple del ajedrez. Son casi triviales, en comparación con la estrategia.

GARRI KASPÁROV

guidamente podríamos anotar jugada a jugada cuál es el mejor movimiento que puede realizarse para cada una. Sin embargo, si calculamos el número de jugadas posibles, nos sale una cifra de alrededor de 10⁵⁶ ceros. Un número descomunal, que es mayor que el total de electrones del universo. Por tanto, solo para guardar los resultados deberíamos tener una memoria mayor que la masa total del universo, lo que evidencia que en este caso sería totalmente imposible simular la inteligencia mediante un diccionario de jugadas-movimientos, a diferencia de lo que sucedía en la habitación china.

A pesar de ello, los ingenieros de IBM presentaron en 1996 a Deep Blue, un supercomputador programado para jugar al ajedrez con gran atino. IBM propuso enfrentar su invento a la inteligencia del campeón mundial de ajedrez del momento, el mítico Garri Kaspárov. Se disputaron un total de seis partidas y el resultado fue de cuatro a dos a favor de Kaspárov. La máquina no ganó al hombre en una primera instancia pero, desde luego, el test de Turing estaba ya superado. En aquella época, Deep Blue podía analizar hasta cien millones de movimientos por segundo.

Sin embargo, el equipo de IBM no se dio por vencido y continuó trabajando en el invento hasta conseguir una segunda versión muy mejorada del Deep Blue original, la Deeper Blue, la cual ya podía analizar doscientos millones de movimientos por segundo. Deeper Blue se enfrentó a Kaspárov en 1997, y esta vez venció al campeón, tras ganar dos partidas de seis y haber conseguido tablas en otras tres. Sorprendentemente, Kaspárov acusó a IBM de hacer trampas, ya que en una de las partidas el

célebre ajedrecista ruso intuyó que la máquina había sido ayudada por un operador humano. La situación fue la siguiente: Kaspárov planteó una jugada en la que, sacrificando un peón, podía armar un contraataque en jugadas subsiguientes. Esta trampa era imposible de detectar por el ordenador, ya que su capacidad de análisis era de un número limitado de jugadas futuras, en las que todavía no se desarrollaba el contraataque del jugador. Sin embargo, Deeper Blue no cayó en la trampa, lo cual levantó las sospechas del campeón ruso. Kaspárov pidió posteriormente los registros por escrito de los procesos de la máquina, a lo cual IBM accedió, aunque finalmente nunca los presentó.

Sea como fuere, Deep Blue marcó un hito en el imaginario popular. Paradójicamente, el gran desarrollo que experimentó la inteligencia artificial en la primera década del siglo XXI pasó desapercibido para el público en general.

¡ELEMENTAL QUERIDO WATSON!

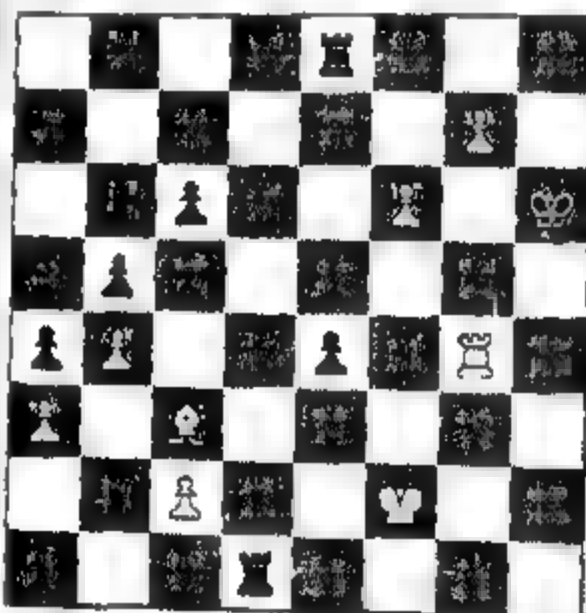
Está claro que tanto Deep Blue como su versión mejorada, Deeper Blue, superaron el test de Turing. Sin embargo, tenemos que reconocer que era un test de Turing un poco especial, ya que el contexto era solo el ajedrez. Por supuesto, aún quedaba por aparecer una máquina que superara esa prueba en toda su extensión, tal y como Turing auguró en la década de 1940: un computador capaz de entender preguntas, procesarlas y responderlas correctamente.

Este hito llegó en febrero de 2011, otra vez de la mano de IBM. Los ingenieros de esta empresa desarrollaron una máquina bautizada Watson y la sometieron a prueba en el concurso televisivo *Jeopardy!*, en una serie de preguntas y respuestas en las que entran en juego conocimientos de todas las ramas del saber, además de numerosos juegos de palabras. La dinámica del juego consiste en que uno de los tres concursantes elige un panel del tablero de juego y se revela una pista en forma de pregunta. A continuación, los concursantes tienen que pulsar un botón. El que lo pulsa primero tiene la oportunidad de dar su respuesta en

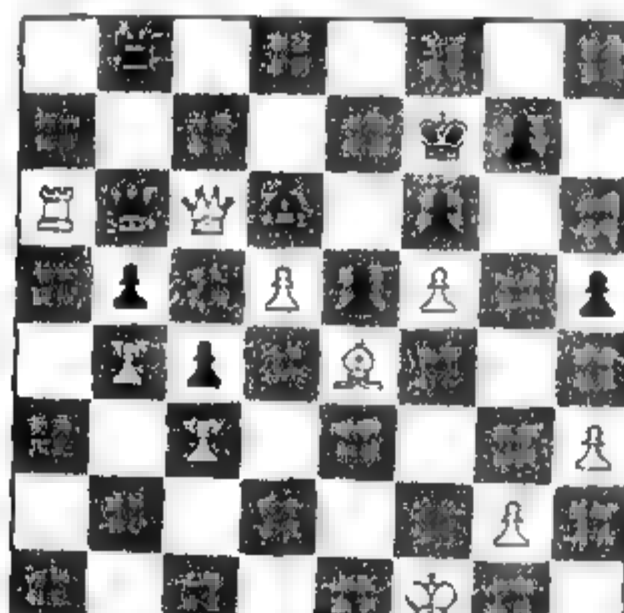
LA PARTIDA DE LA DISCORDIA: EL RELATO DE LOS HECHOS

Para los amantes del ajedrez, se exponen los movimientos de las seis partidas entre Garry Kaspárov y Deeper Blue, la versión mejorada de Deep Blue. La primera de ellas, jugada el 3 de mayo de 1997, se inició con el ataque indio de rey. Kaspárov consiguió ganar a Deeper Blue en 45 movimientos. La segunda partida fue la que desató la tempestad: en ella Kaspárov acusó de hacer trampas al equipo de IBM. Se disputó el 4 de mayo de 1997 y Kaspárov se acabó rindiendo, aunque análisis posteriores indicaron que podía haber logrado el empate. En el tercer enfrentamiento, del 6 de mayo de 1997, Kaspárov utilizó una apertura irregular para poder descolocar a Deeper Blue al tratar de impedir que el computador usara su amplia base de datos de aperturas. A pesar de la táctica, la partida terminó en empate. La cuarta partida, jugada el 7 de mayo de 1997, volvió a terminar en empate y se jugó la defensa Caro-Kann. En el quinto encuentro, celebrado el 10 de mayo de 1997, se repitió el ataque indio de rey y, según los expertos, Deeper Blue hizo una serie de jugadas maestras que bloquearon la ventaja del campeón humano para conseguir unas tablas. Finalmente, la sexta partida entre Deeper Blue y Kaspárov fue ganada por la máquina, el 11 de mayo de 1997. En este caso no hubo polémica y Kaspárov abandonó en el movimiento 19. La conclusión, por tanto, la disputa se saldó por 3 1/2-2 1/2 a favor de Deeper Blue. Los tableros muestran la última jugada de cada una de las seis partidas.

PARTIDA 1



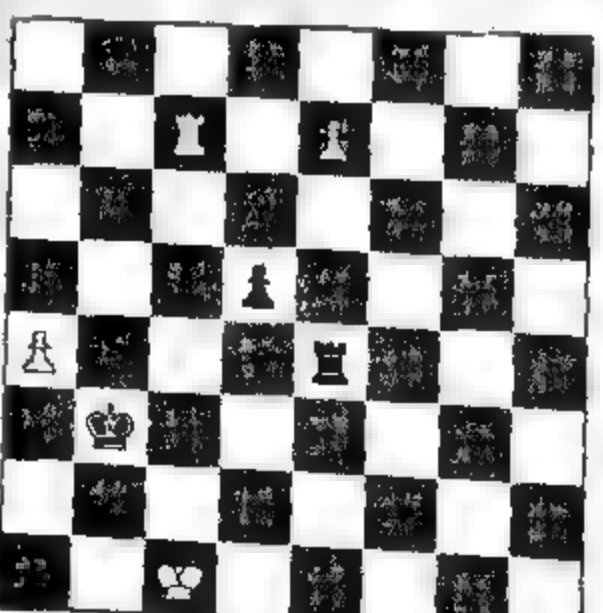
PARTIDA 2



PARTIDA 3



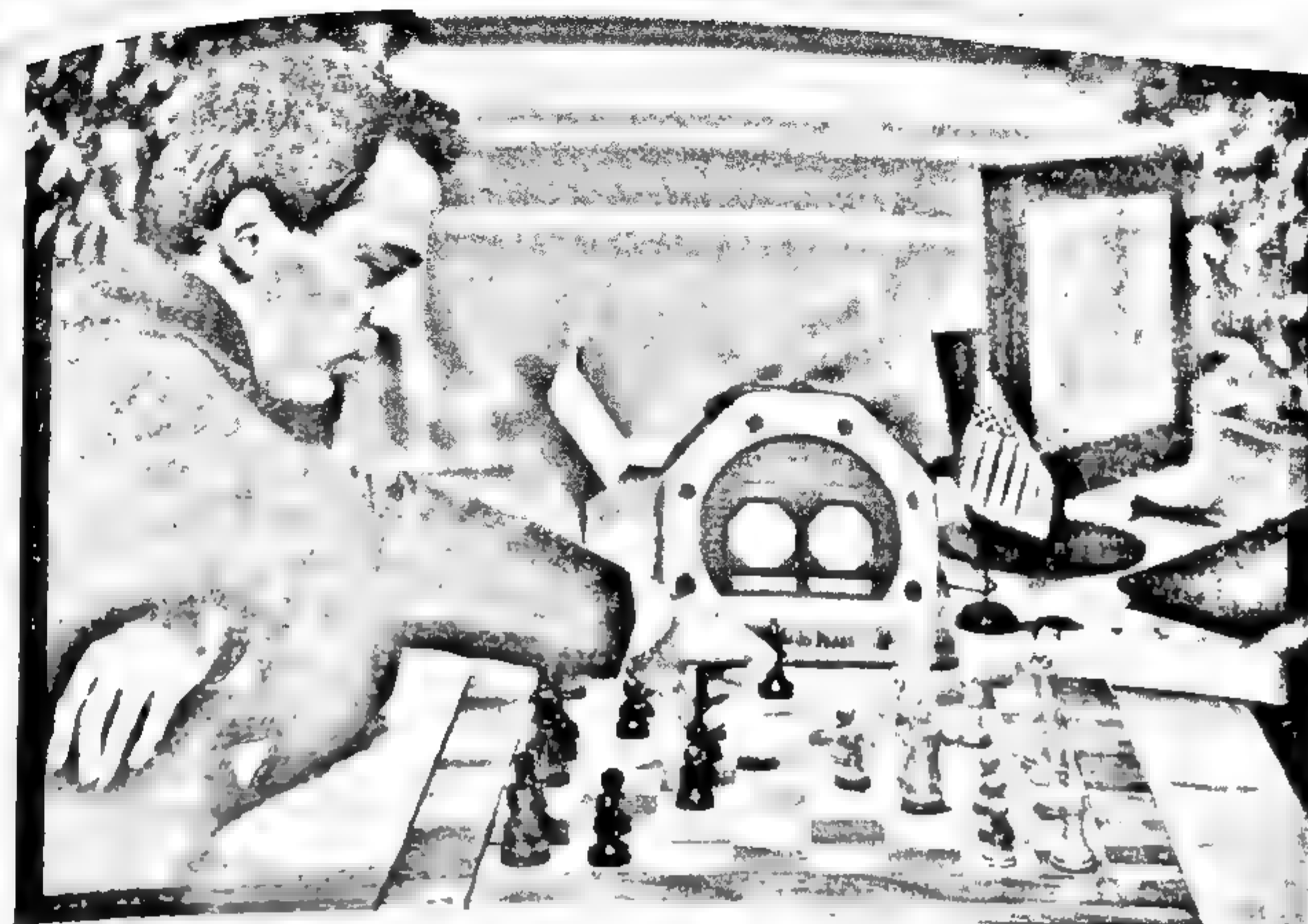
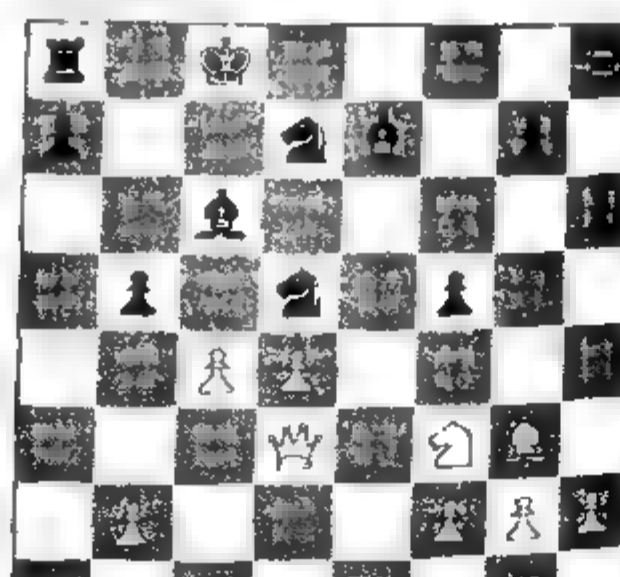
PARTIDA 4



PARTIDA 5



PARTIDA 6



Dos momentos de las partidas que enfrentaron a Garry Kaspárov, campeón del mundo de ajedrez desde 1985 hasta 2000, y al supercomputador Deep Blue.

forma de pregunta, es decir, la pregunta que queda contestada por la respuesta mostrada. Por ejemplo, el panel dice: «Dory». Y el concursante debería responder: «¿Qué está buscando Nemo?», en alusión a la película de Pixar, *Buscando a Dory*. Por supuesto, esta sería una prueba de primer nivel, pero las pruebas se van complicando hasta contener complejos juegos de palabras que involucran conocimientos cruzados de diversas áreas.

Watson fue medido en dos ocasiones ante Ken Jennings y Brad Rutter, dos supercampeones en este juego que habían conseguido ganar varios millones de dólares, arrasando programa tras programa. La gran sorpresa internacional fue que Watson venció a ambos las dos veces con un margen muy significativo, ganando en la contienda un millón de dólares que IBM donó a dos entidades benéficas. Por supuesto, Watson estaba desconectado de internet mientras jugaba, pero esto no impidió que tuviera memorizadas 200 millones de páginas en su memoria RAM, incluyendo, claro está, toda la Wikipedia en inglés, aunque la potencia de un sistema como este no es tener una gran memoria. Al menos, no solo, pues debe saber pasearse por ella en un tiempo récord para crear interrelaciones y poder dar con la respuesta correcta. Y, además, ser capaz de entender el lenguaje natural en el que está escrita la pregunta y poder sintetizar la respuesta con una voz inteligible tanto para la audiencia como para el resto de concursantes. Vamos a transcribir un trozo del programa para tratar de analizar la estrategia de Watson:

Presentador: La excavación de Kathleen Kenyon de esta ciudad mencionada en Josué ha demostrado que sus murallas fueron reparadas en diecisiete ocasiones.

Watson: ¿Qué es Jericó?

Presentador: Esta autora de misterio y su esposo arqueólogo participaron en excavaciones con la esperanza de encontrar la ciudad siria perdida de Arkesh.

Watson: ¿Quién es Agatha Christie?

Presentador: En 1959, en la garganta de Olduvai, ella y el marido Louis encontraron un cráneo de *Australopithecus boisei* de 1,75 millones de años de antigüedad.

Watson: ¿Quién es Mary Leakey?

Presentador: Harriet Boyd Hawes fue la primera mujer que descubrió y excavó un asentamiento minoico en esta isla.

Watson: ¿Qué es Creta?

Presentador: En el monte Carmelo, en Israel, Dorothy Garrod fue la primera que encontró este esqueleto prehistórico fuera de Europa.

Ken Jennings: ¿Qué es el neandertal?

Veamos qué ha pasado aquí. Hay una serie de respuestas que Watson las ve clarísimas, en este caso las cuatro primeras, y es el primero en pulsar el botón para pedir el turno de palabra. Por descontado, dándole a un pulsador, un humano difícilmente puede vencer a una máquina —aunque tenga un dedo robótico, como es el caso de Watson—. Por tanto, en la última pregunta, no es que Watson no quisiera pedir el turno de palabra, sino que deliberadamente cede la palabra al oponente, quizá para no dar una respuesta incorrecta o, tal vez, porque pensaba que su oponente iba a equivocarse.

Evidentemente, Watson cometió algunos errores de bulto a lo largo de los dos programas en los que participó. Por ejemplo, uno de ellos fue repetir una respuesta incorrecta que otro participante había contestado con anterioridad, cosa que hizo porque no escuchó la respuesta de los otros participantes, y no pudo descartar esta respuesta que él tenía como primera opción y proporcionar la segunda que, según sus programadores, tenía como alternativa. Esto nos lleva a concluir que Watson no solo busca la respuesta a una velocidad de vértigo, sino que, además, genera varias respuestas y, después del proceso, las prioriza. En otras palabras, tiene un pensamiento en paralelo, cosa que los humanos somos incapaces de hacer —aunque algunos digan lo contrario—.

Watson cometió además otro error importante que lo delata como no humano. El motivo fue la sintaxis con la que estaba construida la pista, lo que confundió a la máquina, llevándole a pensar que una respuesta era Toronto y no Chicago. La pregunta era: «Su aeropuerto más grande fue nombrado por un héroe de la Segunda Guerra Mundial; su segundo más grande, por una

batalla de la Segunda Guerra Mundial». El problema en esta ocasión, según los ingenieros, fue que Watson no entendió a qué se refería la segunda parte de la pista, es decir, el texto después del punto y coma. Sintácticamente, la segunda proposición debería haberse entendido por sí sola; sin embargo, cualquiera de nosotros —o sea, un humano— hubiera contextualizado y, aunque en la segunda proposición no se habla de ningún aeropuerto, por el contexto de la primera parte de la pista hubiéramos entendido que «el segundo más grande» se refería a un aeropuerto. Pero Watson no lo entendió así, y concluyó que era Chicago. Sea como fuere, su éxito fue abrumador, sorprendiendo no solo a los dos supercampeones, Ken Jennings y Brad Rutter, sino a todo el equipo de ingenieros que lo construyó, como se constata en los vídeos disponibles en internet.

Ahora demos un vistazo a la arquitectura de Watson. Según IBM, es el resultado de la aplicación de tecnologías avanzadas diseñadas para el procesamiento de lenguajes naturales, la recuperación de información, la representación del conocimiento, el razonamiento automático y el aprendizaje automático al campo abierto de búsquedas de respuestas, que es construido con la tecnología DeepQA de IBM para la generación de hipótesis, la recopilación de pruebas masivas, el análisis y la calificación. DeepQA es un software que usa más de cien técnicas diferentes para analizar el lenguaje natural, identificar fuentes, encontrar y generar hipótesis, buscar y puntuar evidencias, combinar y clasificar hipótesis. De hecho, DeepQA usa técnicas de *big data* que veremos un poco más adelante.

Por lo que respecta al hardware, el ordenador que soporta a Watson es un supercomputador de dimensiones considerables. Concretamente, está compuesto por 90 servidores IBM Power 750, cada uno de ellos de ocho núcleos. En total, Watson dispone de 16 terabytes (TB) de memoria RAM, lo que le permite albergar en memoria la ingente cantidad de información que antes hemos mencionado: unos 200 millones de páginas. En su conjunto, es capaz de procesar el equivalente a un millón de libros por segundo. Por tanto, ya sea por su inteligencia o por su velocidad, Watson se convirtió en un contrincante invencible en este tipo de juegos.

Pero ¿sirvió a la humanidad para algo más que impresionarnos? La respuesta, obviamente, es afirmativa. En el terreno del desarrollo de la inteligencia artificial, el equipo que lo construyó consiguió avanzar enormemente el estado de la técnica, y a nivel comercial, IBM aún está sacando gran provecho de aquel invento: en concreto, está comercializando versiones «light» de Watson para ayudar a los clientes de grandes empresas como bancos o telefónicas a responder a sus preguntas y dudas sobre el servicio. IBM también ha lanzado una versión de Watson un 75% menor y un 25% más rápida para poder responder a los usuarios desde sus teléfonos móviles. Todo esto aporta pingües beneficios a la compañía, pero lo que realmente representa un grandísimo avance para la humanidad es que Watson ya está siendo usado como médico de cabecera con una fiabilidad bastante mayor que la de un médico humano.

Cuando un paciente va a visitar a un médico con una rara enfermedad, el profesional no solo tiene que contrastar un complejo número de resultados analíticos —que muchas veces son difíciles de entender para un mismo profesional— sino que debe estar completamente al día de las miles de enfermedades que se van describiendo, de sus tratamientos y, por supuesto, de los centenares de descubrimientos científicos que a diario se publican en las revistas de referencia. Además de todo ello, el médico debe eliminar, en la medida de lo posible, sus sentimientos de la ecuación y tratar de guiarse por su raciocinio sin dejarse influenciar por ningún tipo de sesgo. En definitiva, una tarea perfecta para Watson. Pero ¿están los pacientes listos para ser diagnosticados por una máquina? Y lo que es peor, ¿están los médicos listos para ser sustituidos por ella? La respuesta la tendremos en los próximos años.

LA CLAVE ESTÁ EN LA NUBE

Aunque no se trate específicamente de inteligencia artificial, es relevante hablar de un nuevo concepto que ha aparecido en los últimos años y que ha revolucionado los sistemas de computación: se trata del *cloud computing* o la computación en la nube,

otro ejemplo de cómo el avance de la técnica es siempre concomitante al avance de la capacidad de computación.

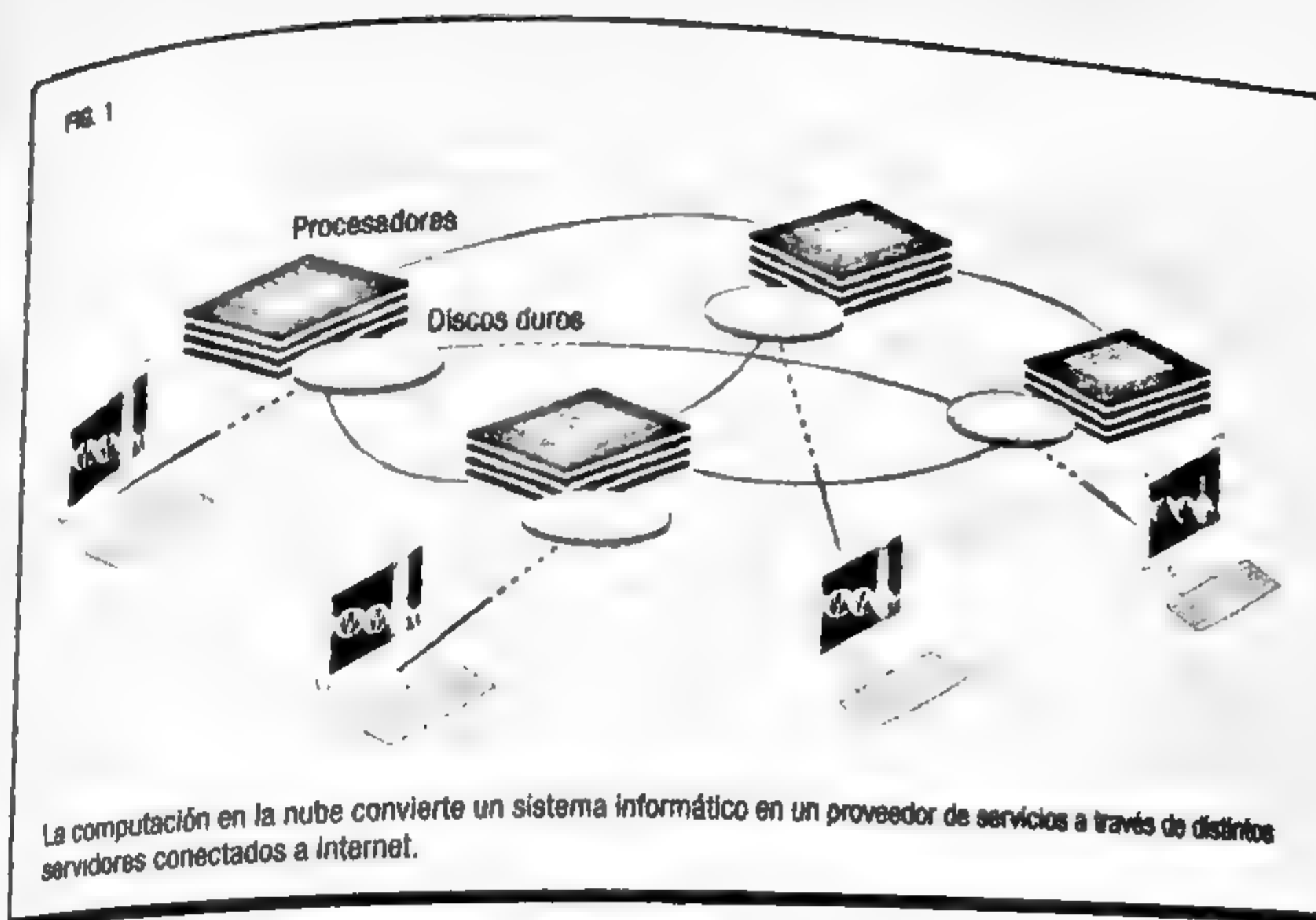
El precursor de la computación en la nube fue lo conocido como *grid computing* o «computación en malla». El *grid computing* fue un concepto acuñado a principios del siglo XXI que intentaba imitar los sistemas de cogeneración eléctrica en *grid*. Como sabemos, en muchos países desarrollados la cogeneración eléctrica está completamente distribuida por el territorio. En vez

La computación ya no va más sobre computadores. Va sobre la vida.

NICHOLAS NEGROPONTE

de tener grandes centrales eléctricas, estos países han tomado la decisión estratégica de tener centenares o miles de pequeñas centrales distribuidas por toda su geografía. Aunque esto pueda parecer ineficiente, la ventaja de este sistema en caso de accidente, conflicto armado o acto terrorista es enorme, ya que si se destruye una de estas centrales, el sistema no se ve prácticamente afectado. Además, el sistema incentiva que sean los grandes consumidores, como por ejemplo las grandes fábricas, los propietarios de estas centrales, de tal forma que, mientras la fábrica está activa, seguramente la potencia eléctrica necesaria es mayor a la generada por la propia central, así que esta consume un extra de energía de la red. Sin embargo, cuando la fábrica está parada, como sucede durante la noche, la central eléctrica inyecta la energía sobrante en la red para su consumo en otros lugares, como los hogares. A final de mes, las empresas de distribución eléctrica pueden calcular el balance entre lo consumido de la red eléctrica y lo inyectado, y saldan la cuenta con el propietario de cada central. En definitiva, un sistema simple, eficiente y seguro.

Por tanto ¿por qué no copiar el mismo concepto en la computación? Si varias universidades o centros de investigación poseen grandes supercomputadores, estos pueden ser unidos mediante una red telemática —internet— (figura 1) que permita a los investigadores exportar sus cálculos de un supercomputador a otro, en caso de que el suyo propio esté saturado. De la misma forma que en el caso de la energía, al final del periodo tarifario un operador neutral puede contabilizar los ciclos de cálculo que cada



universidad ha cedido o tomado prestados de otros centros y generar una factura entre la red de universidades para compensar el balance.

Este sistema se puso en marcha en diversas partes del mundo, aunque de forma limitada y sin demasiado éxito comercial. Sin embargo, en paralelo a todo esto y al desarrollo tecnológico que permitió el *grid computing*, la empresa Amazon desarrolló una necesidad que exigía respuesta inmediata. Como es bien sabido, Amazon se inició como una empresa de venta *online* de libros, aunque hoy en día haya evolucionado a un inmenso hipermercado en línea. En aquellos inicios, la gran mayoría de sus ventas se producía en época navideña. Amazon no quería desaprovechar la campaña y se veía obligada a dimensionar sus servidores durante las fechas punta, de manera que los usuarios pudieran hacer tantas compras como quisieran en esos días sin que los servidores se colapsaran. Esto provocó que la empresa atesorara una gran capacidad de cálculo que, en los once meses restantes, estaba prácticamente inactiva.

Para poder rentabilizar esta extraordinaria potencia de cálculo el resto de los meses del año, Amazon ideó el concepto de *cloud computing*: un sistema que permitía alquilar por precios realmente competitivos tanta potencia de cálculo como se quisiera, aunque el supercomputador estuviera ubicado en la otra parte del mundo. La expresión «computación en la nube» es, por ende, una metáfora que simboliza que determinados cálculos han ido a ejecutarse a una nube abstracta que nadie sabe realmente dónde está ni qué aspecto tiene —haciendo referencia a la configuración exacta de los servidores—. Al final de esos cálculos, Amazon factura en función del uso hecho de estos recursos computacionales. Curiosamente, y según se cuenta, hoy en día Amazon gana más dinero gracias a sus servicios de computación en la nube que por la venta de libros, razón por la cual otras empresas informáticas como Google o Microsoft han desarrollado sus propios sistemas de computación en la nube, abiertos al público y también a precios realmente competitivos. En definitiva, hoy en día cualquiera puede ser el propietario de un gran supercomputador... aunque sea solo por unos pocos segundos.

La segunda gran innovación introducida en este ámbito es la virtualización. Cuando hacemos cálculos en la nube existe un problema obvio: no tenemos ni idea de en qué ordenador nos va a tocar hacer el cálculo ni qué configuración tendrá. Como es fácil de imaginar, existen infinitas configuraciones posibles de un computador: número y tipo de procesadores, memoria RAM, espacio en disco duro, tipo de interconexión entre los nodos, etc. En el proceso de virtualización, el usuario especifica el tipo concreto de configuración que desea y automáticamente se construye una máquina virtual con la configuración especificada que se ejecuta sobre una máquina física cualquiera. Si la máquina virtual no ocupa una máquina física completa —por ejemplo, en el caso de que lo demandado por el usuario ocupe menos espacio que la máquina donde va a ejecutarse—, varios usuarios pueden compartir, sin saberlo, dicha máquina sin que haya ningún tipo de visibilidad entre ellos, ya que es materialmente imposible salirse de un entorno virtual para entrar en el siguiente.

Por tanto, la virtualización, junto con la computación en la nube, no solo permiten a cualquiera tener un supercomputador en cualquier momento por un precio muy competitivo, sino que, además, le permite disponer del supercomputador preciso que necesita para cumplir sus objetivos de cálculo. Estos dos conceptos han experimentado una gran eclosión en la segunda década del siglo XXI, permitiendo la «viralización» de muchas aplicaciones web que sin ellos no hubieran sido posibles. Hoy, si una aplicación tiene mucho éxito, se puede ampliar la máquina sobre la que corre en tiempo real simplemente añadiendo más unidades virtuales de la nube al servidor que da soporte a la aplicación en cuestión. Y esto mismo se puede extender a los sistemas de inteligencia artificial. Los ingenieros tienen ahora la posibilidad de disponer tan solo de un nodo central ejecutándose de forma permanente por muy pocos céntimos/hora y, en caso de que surja una necesidad potente de cálculo, simplemente se activarán más nodos virtuales que añadirán a ese supercomputador la suficiente potencia de cálculo para poder dar respuesta inmediata a los requisitos demandados.

Big data en el día a día

La inteligencia artificial ha hecho posible, entre otras muchas cosas, que nuestro ordenador nos brinde lo que necesitamos incluso antes de acabar de teclearlo. Las máquinas, que pueden saberlo todo sobre nosotros, ya son capaces de adelantarse a nuestras necesidades. Bienvenidos a la nueva era.

Los datos masivos o *big data* constituyen la piedra filosofal gracias a la cual las máquinas acceden a todo el universo de costumbres e intereses de los humanos, lo que les permite anticiparse a nuestras necesidades. Pero ¿de qué trata el *big data* exactamente? A diario oímos hablar de ello en la televisión, en los periódicos o en la calle. Y es que, en la era actual, este término va ligado a un sinfín de aplicaciones: desde saber cómo dirigir y planificar las campañas electorales para sacarles el máximo rédito hasta generar ofertas hechas a medida para satisfacer nuestras futuras demandas, pasando por la producción de contenidos multimedia para la televisión, en función de cuáles sean nuestros gustos. Este concepto, que afronta cómo manejar el aluvión intangible de datos que se genera a cada segundo, se sustenta sobre cinco pilares, conocidos como las cinco «V» del *big data*: Volumen, Velocidad, Variedad, Veracidad y Valor.

Empecemos por la primera V, la que hace referencia a «volumen». Con el advenimiento y generalización de las redes sociales, el volumen de información que la humanidad genera a cada instante es astronómico. Solo por poner algunos ejemplos, en noviembre de 2016, mientras este libro estaba siendo escrito,

cada segundo se publicaban en todo el mundo unos 6000 tuits, se realizaban más de 40 000 búsquedas en Google, se conectaban 12 000 personas a Facebook y se enviaban más de dos millones de correos electrónicos, todo en números redondos y aproximados. Unos datos que, bien procesados, son una mina de información y un valiosísimo recurso de conocimiento. Pero, además, está la información que, ya sea públicamente o a través de nuestras transacciones comerciales, está disponible sobre todos nosotros; no solo son los *likes* de Facebook, los tuits que realizamos o los vídeos que miramos en YouTube; hay mucho más de lo que ni siquiera somos conscientes. Por ejemplo, los pagos que realizamos con nuestra tarjeta de crédito, nuestras rutas geográficas más comunes —recordemos que hoy quien más y quien menos tiene un teléfono móvil con GPS— o los flujos de información que generamos, ya sea mediante conferencias telefónicas, WhatsApp o cualquier otra herramienta sustitutiva.

Toda esa información, analizada con un potente y eficaz motor de inteligencia artificial, permite conocer incluso mejor que nosotros mismos nuestros hábitos y costumbres. Se cuenta que el supermercado estadounidense Target se dedicaba a enviar a sus clientas cupones de descuento para productos premamá y para el bebé cuando detectaba mediante patrones de comportamiento (cuando estas compraban productos como cremas, suplementos vitamínicos y otros) que podían estar embarazadas. La anécdota, que si no es cierta bien podría serlo, narra que un padre acudió al centro furioso para pedir explicaciones sobre cómo era posible que mandaran esos productos a su hija aún adolescente. Y es que esta, como se supo enseguida, sí estaba embarazada pero todavía no había comunicado a su familia la «feliz» noticia...

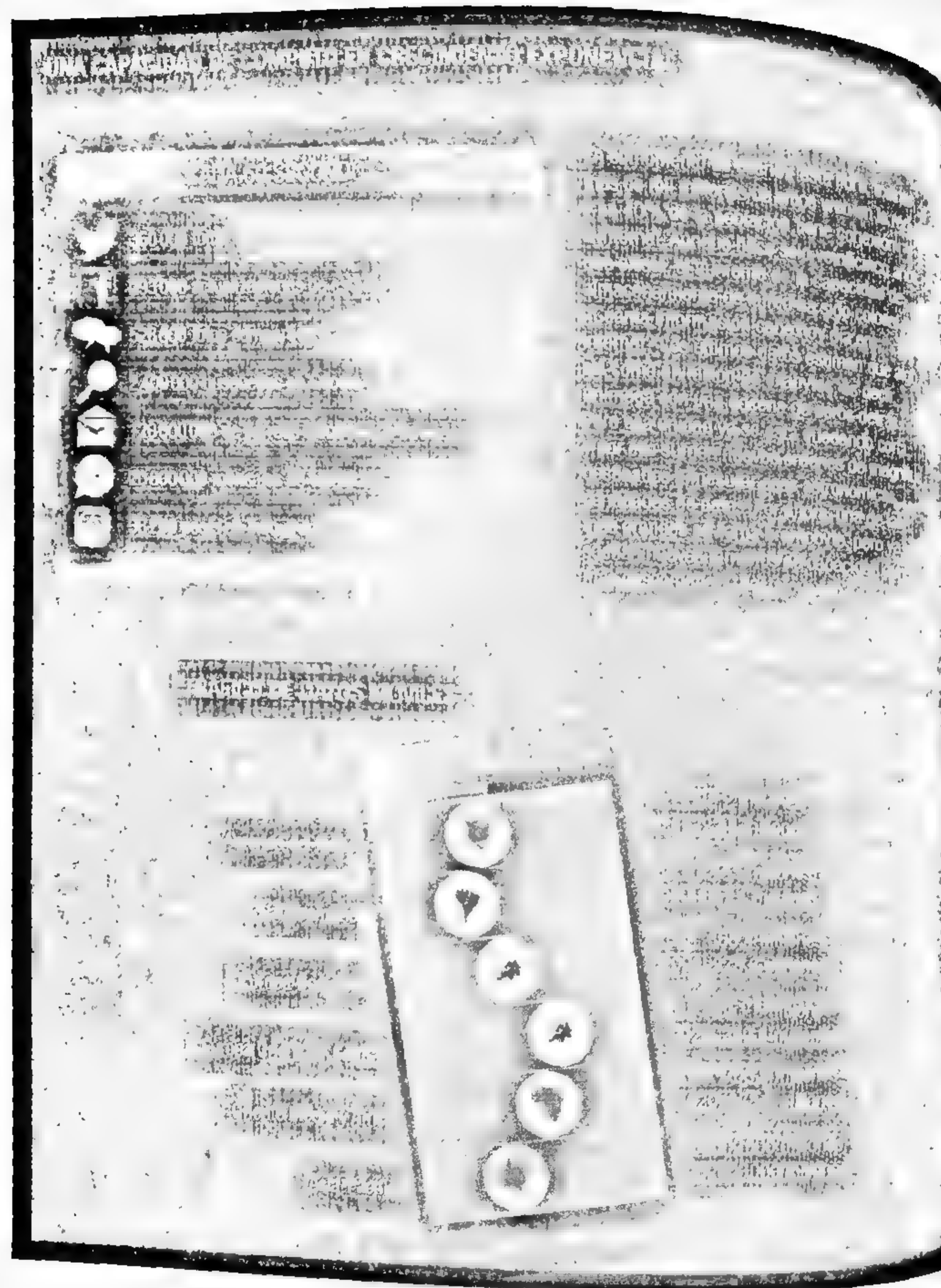
La segunda V hace referencia a la «velocidad». Y es que si queremos saber, por ejemplo, el tiempo que va a hacer mañana, es evidente que no resulta operativo realizar un complejo cálculo físico de las condiciones atmosféricas que nos lleve media semana, ya que para cuando tengamos el pronóstico, ya habrá pasado el periodo objeto de la predicción. Algo parecido sucede con el *big data*: las predicciones de las necesidades deben ser

calculadas antes de que estas estén cubiertas. Por tanto, la velocidad, tanto en el proceso de datos como en la extracción de conclusiones que se deriven, es crucial para que todo el proceso tenga una utilidad práctica. Un ejemplo: la multinacional Telefónica, S.A. publicó hace unos años su proyecto LUCA, realizado por su departamento de I+D, en el que usaba el *big data* para detectar en tiempo real las réplicas de un potente terremoto que hubo en Oaxaca, México, en el año 2012 y poder alertar así a los servicios de emergencia. El conocimiento del detalle de las réplicas se sacaba del flujo de llamadas que se realizaban cada segundo: si, de repente, en un barrio aumentan las llamadas telefónicas un 5 000% por encima de la media, es que está teniendo lugar algún suceso extraordinario que los abonados quieren comunicar a sus allegados. En este caso, se trataba de una réplica del terremoto.

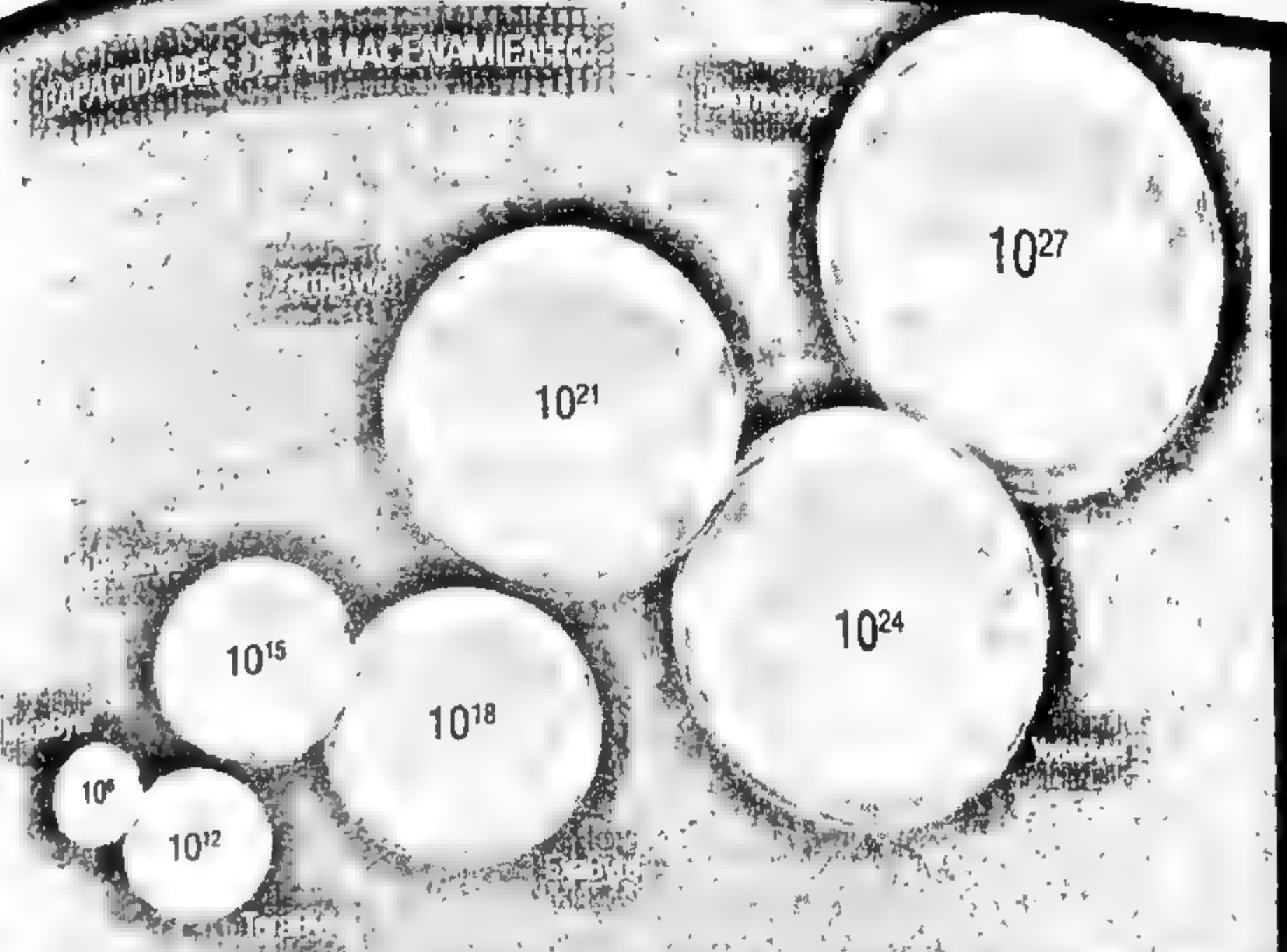
La tercera V está vinculada al concepto de «variedad». Sabemos que el gran volumen de datos que la humanidad genera a cada instante es tremendo. Esto no sería un gran problema si no fuera porque la variedad y heterogeneidad de estos datos también es enorme. ¿Cómo se puede extraer conocimiento de una serie de tuits combinados con un grupo de búsquedas por Google, un puñado de fotos de Instagram y unos cuantos millones de *likes* en Facebook? De esto trata la V de variedad. Cuando los ingenieros implementan sistemas de inteligencia artificial basados en el *big data*, la diversidad de los datos recibidos es un aspecto crucial que hay que tener en cuenta. El reto de la inteligencia artificial es poder extraer conclusiones no solo de los textos escritos en lenguaje natural, sino también de fotos, vídeos, importes de compras en TPV (Terminal Punto de Venta), llamadas telefónicas y un largo etcétera. Los tecnólogos han desarrollado al respecto el concepto denominado *bases de datos no estructuradas*, en contraposición a las clásicas bases de datos relacionales, donde toda la información está tabulada siguiendo unos mismos y estrictos patrones. En las bases de

Sin el análisis del *big data*, las compañías están sordas y ciegas, corriendo por la web como un cervatillo por una autopista.

GEOFFREY MOORE



CAPACIDAD DE ALMACENAMIENTO



datos no estructuradas, la información se almacena «de cualquier manera» y luego es responsabilidad de los algoritmos de inteligencia artificial poder explorar e interrelacionar esta información y sacar las conclusiones pertinentes.

La cuarta V pertenece al término «veracidad» y probablemente es la más controvertida. Sin duda es incuestionable que se genera un ingente volumen de información, pero puede suceder que esta sea falsa y que, por consiguiente, confunda al sistema. Esto implica que los sistemas de inteligencia artificial basados en *big data* deben tener la destreza suficiente como para discernir entre la información veraz y el ruido de fondo. La polémica surge cuando algunos autores afirman que, en caso de generarse alguna información falsa o inexacta, esta se produce en cantidades tan ínfimas que su significación estadística es irrelevante. Por ejemplo, ¿cómo se pueden crear registros de compra falsos con un volumen suficiente como para que confunda a los registros reales?

Por último, está la V de «valor», la V que ha convencido de forma definitiva a todas las grandes corporaciones de que deben volcar sus esfuerzos en el *big data*. Es fácil adivinar el considerable valor que se puede extraer de los datos masivos generados por los ciudadanos. Las grandes corporaciones, al poder adelantarse a nuestras necesidades y ser capaces de saber mucho sobre nosotros y sobre nuestras costumbres, crean nuevas líneas de negocio que les generan pingües beneficios, mientras que al ciudadano le hacen la vida un poco más agradable, ahorrándole tiempo y, quizá, algún disgusto. Solo por poner un par de ejemplos: si hoy he tenido un día muy ajetreado y he dejado rastro mediante datos generados que responden a este patrón de estrés, cuando llegue a casa, mi televisor podría iniciarse con una música relajante sin que yo tenga que programarlo ni darle ninguna indicación. O si en mi entorno hay una epidemia de gripe —y en esto de las epidemias, Google es muy bueno gracias a los cambios de los patrones de búsqueda de los usuarios— puedo recibir una notificación automática que me advierta de ello y me facilite cita con el médico para que me administre la vacuna correspondiente.

ALMACENAMIENTO, PROCESAMIENTO Y COMUNICACIONES: LOS TRES PILARES DEL *BIG DATA*

Analicemos ahora la vertiente más tecnológica y menos conceptual del *big data*, es decir, la infraestructura sobre la que necesariamente se tiene que sustentar. Cualquier sistema de *big data* debe tener bien resueltos tres aspectos sin los cuales sería imposible un buen funcionamiento: el almacenamiento, el procesamiento y las comunicaciones. En cuanto al almacenamiento, ya hemos dado unas pinceladas en los párrafos anteriores. Solo por dar una cifra: un solo motor de avión genera 10 TB de datos cada 30 minutos gracias a todos los sensores que lleva incorporados. Por tanto, si alguien se dispone a aplicar técnicas de *big data* sobre este tipo de información, obviamente necesita grandísimos espacios donde ir acumulándola y, además, ¡en tiempo real! Y es que en caso de que el sistema de almacenamiento, por lo que fuere, se colapsara por un instante, se perderían miles de datos que no quedarían grabados.

Para solucionar este problema existe una gran variedad de tecnologías que permiten tanto el almacenamiento en sistemas hardware distribuidos como la administración de datos no estructurados y tan poco amigos de las bases de datos tradicionales, tal y como comentábamos en la V de variedad.

El segundo aspecto que debe tenerse en cuenta es el procesamiento. En este punto, justamente, es donde interviene la inteligencia artificial. Por ello hablaremos del *deep learning*, que es una técnica de inteligencia artificial basada en un enfoque conexionista y ampliamente usada en entornos de *big data*. Sin embargo, antes de aplicar estas técnicas de procesamiento de datos que tienen la finalidad de proporcionar conocimiento a partir de los datos, estos deben ser organizados de modo que puedan ser manejables de una u otra manera. Por ello se ha desarrollado la técnica del *map-reduce*, popularizada, precisamente, por la compañía Google.

Esta técnica consiste en un modelo de programación que, mediante computación paralela, permite disgregar las diferentes partes de un problema —lo que se denomina *map*— entre dife-

rentes nodos de cálculo para, una vez resuelto el cálculo individual de cada una de las partes, integrar el resultado en un solo dato —lo que se conoce como *reduce*—. En definitiva, la función *map* se encarga del mapeo de los datos recopilados entre los diferentes nodos de cálculo, y la función *reduce*, de combinar los múltiples resultados individuales obtenidos en un único resultado, teniendo necesariamente que simplificar —o reducir— la complejidad del conjunto de resultados.

La arquitectura típica en un esquema de *map-reduce* es un computador con múltiples nodos donde uno de ellos es el «maestro» que se encarga de trocear el problema y enviarlo a los nodos de cálculo o «trabajadores»; cuando estos terminan, procede a fusionar el resultado. Los nodos «trabajadores» son los que realmente implementan la lógica del procesamiento de datos, es decir, los algoritmos de inteligencia artificial que permiten extraer las conclusiones, información y conocimiento de los datos que reciben de entrada.

Finalmente, el tercer aspecto que se ha de tener en cuenta en un entorno de *big data* son las comunicaciones. No hay duda de que la infraestructura de telecomunicaciones que debe existir para dar soporte a un entorno de datos masivos tiene que poder proporcionar unas velocidades de transferencia de vértigo para dar respuesta en tiempo real a las necesidades del usuario. Las infraestructuras de fibra óptica que se han generalizado en los últimos tiempos son suficientes. Sin embargo, no solo la velocidad de transferencia es crucial, también lo es la latencia, o lo que es lo mismo, el tiempo que transcurre desde que un bit sale del origen hasta que llega al destino. Como ejemplo podemos decir que nuestra «velocidad de transferencia» es mayor si vamos en avión que en coche. Sin embargo, nuestra «latencia» va a ser mucho menor si vamos en coche de nuestra ciudad a un pueblo cercano que de Tokio a Nueva York en avión supersónico. De aquí la importancia del *cloud computing* que vimos en el capítulo anterior y de la necesidad de diseñar de forma óptima la distribución geográfica de las nubes de servidores que componen una nube.

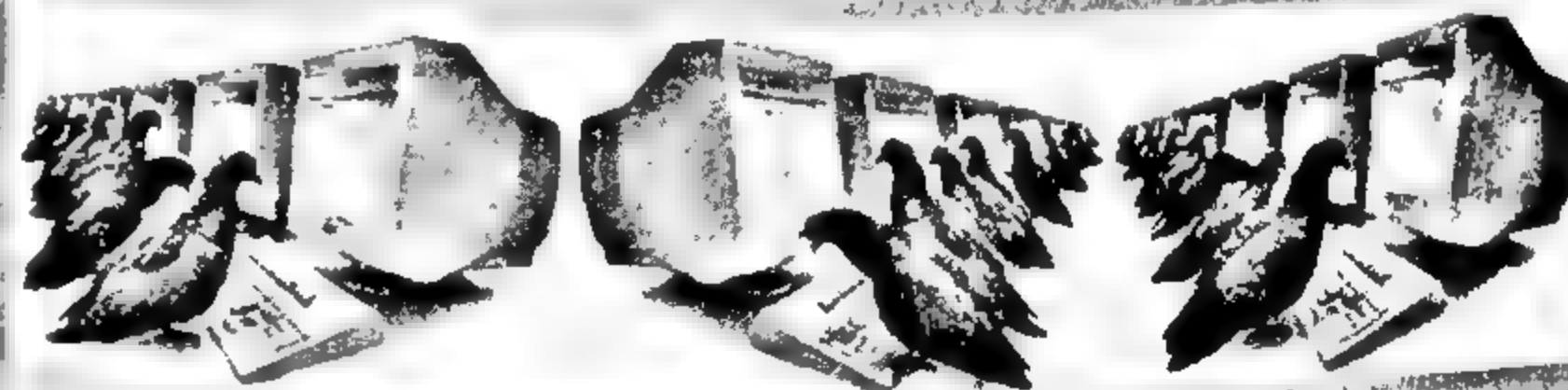
Contaremos un caso práctico a modo de conclusión. El *big data* tiene aplicaciones en todos los ámbitos de la economía,

LA INOCENTADA DEL PIGEON RANKING

Google presume de una gran reputación en el mundo del *big data*. De hecho, su santo grial es el algoritmo usado para confeccionar los rankings de las búsquedas que los internautas de todo el mundo realizan millones de veces por segundo. Tanto interés despierta este algoritmo y tantas presiones había recibido Google para publicarlo que la mañana del 1 de abril de 2002 la empresa decidió poner un enlace desde su página principal donde, afirmó, explicaba su algoritmo de ranking.

Palomas superdotadas

El algoritmo en cuestión se denomina *Pigeon Ranking* («ranking de palomas») y su extraño nombre no era para nada un capricho. Google disponía de unas instalaciones llenas de PC —siglas en inglés que, en este caso, en lugar de significar *personal computer*, se atribuían a *pigeon cluster*, o lo que es lo mismo, «agrupación de palomas»— donde cada una de las aves disponía de una pantalla y un teclado. Se contaba que, cada vez que un usuario realizaba una búsqueda en Google, cada una de las webs que coincidían con dicha búsqueda era mostrada a una de las palomas que, acto seguido, empezaba a picotear en su teclado. A continuación, las webs eran ordenadas según el número de picotazos que había dado su paloma correspondiente. En el mismo artículo, Google explicaba cómo manipulaban a las palomas, cómo vivían y cómo eran reclutadas. También se decía que habían probado con otro tipo de aves, como las gallinas o distintas especies de aves de presa, pero que las palomas se habían revelado como las más inteligentes y adecuadas para la tarea. El artículo incluso afirmaba que, si bien era cierto que por el momento ninguna paloma había llegado a ser miembro del tribunal constitucional, si se había probado su eficacia como controladores aéreos y árbitros de fútbol. Muchos técnicos del sector no daban crédito a lo que estaban leyendo, hasta que la gente empezó a caer en la cuenta de que el día 1 de abril es, en Estados Unidos y muchos otros países, el denominado *April's Fools' Day*, algo parecido al día de los Santos Inocentes que se celebra en algunos países de habla hispana, en el que la gente suele gastar inocentadas durante toda la jornada. Aquel día, y por un momento, alguien pensó que Google había publicado su gran secreto en cuanto al procesamiento de *big data* se refiere. Años después, en 2014, Google publicó su algoritmo de inteligencia artificial de búsqueda local Google Pigeon. Esta vez, sin embargo, nada tenía que ver con la paloma ni con ninguna otra ave columbiforme.



Google explicó en su día que las palomas eran entrenadas siguiendo las pautas del reputado psicólogo B.F. Skinner, quien realmente enseñó a algunas de ellas a hacer cosas increíbles, como jugar a ping-pong.

como las distintas operaciones, las finanzas, la seguridad, la salud o el marketing. Existe un caso archiconocido y relacionado con este último que viene al caso: el sistema de recomendaciones de Amazon. Este sistema se puede experimentar cada vez que se busca un determinado objeto en esta plataforma, ya sea un usuario registrado o no, en la sección *Customers who bought this item also bought...* («Los clientes que compraron este objeto también compraron...»). Aunque puede parecer trivial, la complejidad que se esconde tras esta simple idea es notable, hasta el punto de que esta herramienta está basada en técnicas clasificadas dentro de la inteligencia artificial y van bastante más allá de explorar simplemente qué hay en las cestas de la compra de otros usuarios que adquirieron el mismo objeto que me interesa a mí.

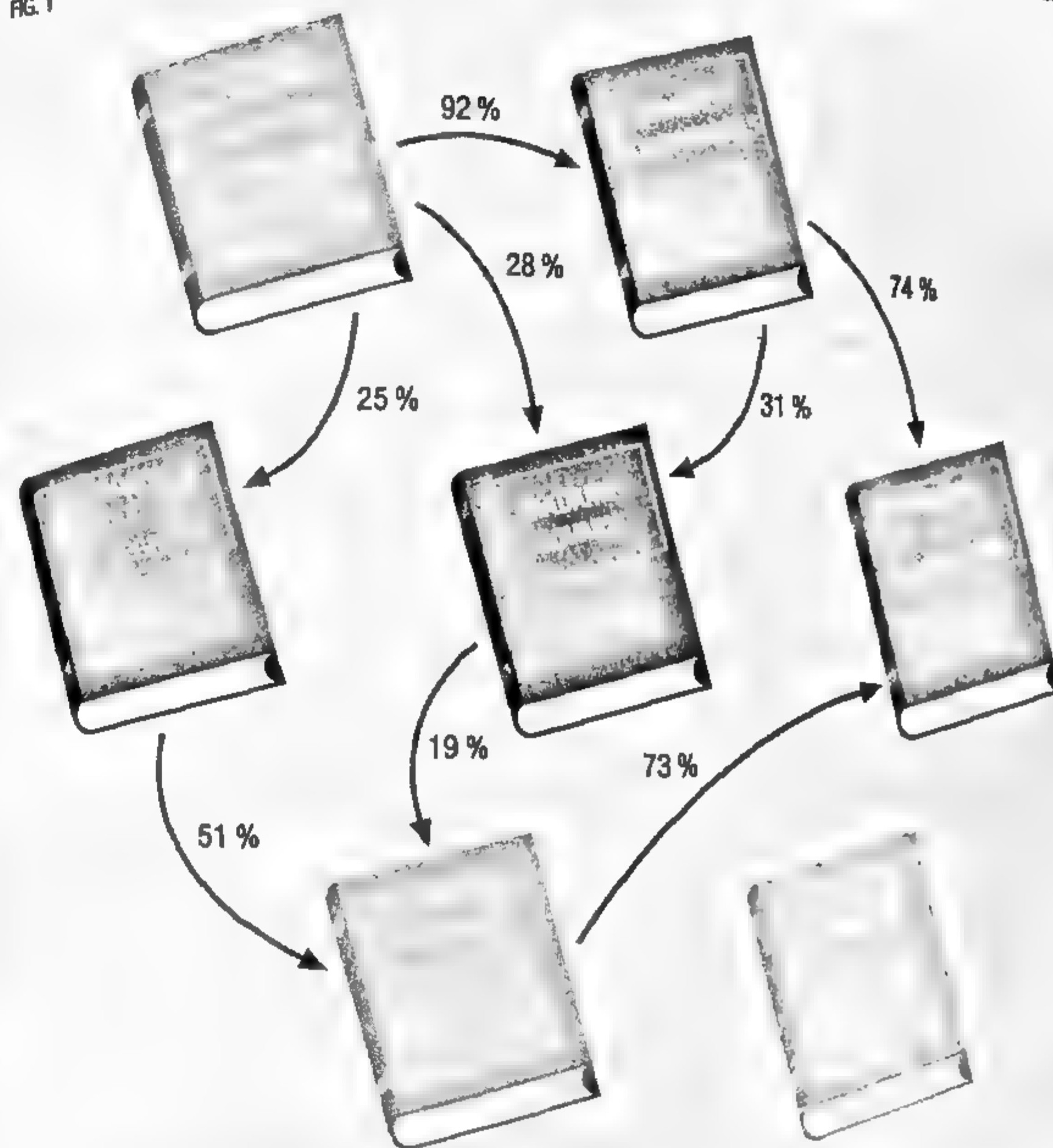
La tecnología que se suele usar para aproximar este tipo de problemas es lo que se conoce como *redes bayesianas*, y el mayor centro de investigación mundial especializado en esta herramienta es la división de Microsoft dedicada a la investigación, el Microsoft Research Institute, donde se estudia la aplicabilidad de esta técnica no solo para el marketing *online*, sino también para otros aspectos, como la adaptación automática de la interfaz de usuario de Windows a cada cliente según su manera de trabajar o sus preferencias.

La idea que hay detrás de una red bayesiana es que hay cadenas de eventos que suelen sucederse, las cuales pueden compartir probabilidades con otras cadenas de eventos —por eso se llaman redes, ya que son cadenas de probabilidades entrecruzadas—. Veamos un ejemplo sobre la compra de libros (figura 1).

En esta red podemos ver que el 92% de los clientes que compraron *Orgullo y Prejuicio* también adquirieron *Cumbres Borrascosas*. En cambio, ninguno de ellos compró *El Capital*, y, por tanto, no se establece ninguna relación entre estos dos libros. Por eso, si el sistema detecta que un cliente acaba de comprar *Orgullo y Prejuicio* y ahora está buscando información sobre *Cumbres Borrascosas*, en el apartado de recomendaciones ya estará apareciendo *Madame Bovary* y *Sentido y Sensibilidad*, ya que un porcentaje muy significativo de la gente que com-

pró *Orgullo y Prejuicio* y luego *Cumbres Borrascosas*, también se hizo con estos dos libros. Con todo esto, el sistema trata de aumentar sus ventas mediante una campaña de marketing individualizada a cada cliente y automatizada para anunciar productos que quizá el cliente no sabía que existían. Gracias a la gran

FIG. 1



Las redes bayesianas presentan un conjunto de variables aleatorias que se influyen entre sí. Aplicadas a la compra *online*, como en el caso de los libros, ayuda a definir y acotar los gustos del cliente, comparándolos con las tendencias generales.

cantidad de información de compradores previos que posee el sistema, este ha podido establecer una red de relaciones causales que usa para hacer nuevas recomendaciones.

Además, el sistema también sabe que es perder el tiempo publicitar el clásico de Karl Marx a un cliente que está comprando

Había 5 exabytes de información creadas desde el alba de la civilización hasta 2003. Pero ahora esto se crea cada dos días.

ERIC SCHMIDT (2010)

clásicos románticos, que es lo que sucedería si hiciéramos una campaña de marketing tradicional. Por ejemplo, en una campaña de marketing clásica, se anunciaría la nueva edición de *El Capital* en el programa temático sobre libros y lectura que se emite en el canal cultural a las 23:00 h. Sin em-

bargo, aunque se habría escogido el programa y la franja horaria donde el público tendría más probabilidad de estar interesado en el producto, seguiríamos teniendo un gran número de telespectadores aficionados a la novela romántica para los que este anuncio no estaría surgiendo efecto alguno, con la pérdida monetaria que eso supone para la empresa anunciante. Para un canal de marketing estático, como la televisión, la radio o los carteles en las calles, es imposible que el anunciante conozca el perfil individualizado de los oyentes del momento y, aunque lo supiera, carece de las herramientas necesarias para adaptar el anuncio emitido a cada oyente. Con el *big data*, sin embargo, el paradigma ha cambiado completamente.

UNA RED DESCOMUNAL DE MILLONES DE NEURONAS: EL DEEP LEARNING

Ya hemos comentado que el enfoque *bottom-up* o modelo conexionista ha ido sufriendo grandes altibajos en la historia de la inteligencia artificial. El último fue provocado por el desarrollo de las máquinas de soporte vectorial. Las redes neuronales adolecen de dos problemas principales. El primero es que su funcionamiento, aunque sea correcto, es incomprensible para la mente humana. El segundo es que tienen tantos parámetros que

es materialmente imposible probar cuál es la mejor configuración para resolver un determinado problema, ya sea el número de neuronas, la función de transferencia o el número de capas ocultas. Por esto Vladimir Vapnik y sus colaboradores propusieron las máquinas de soporte vectorial, ya que solucionan ambos problemas, además de ser un sistema computacionalmente mucho menos costoso que el de una red neuronal.

Sin embargo, con el exponencial desarrollo de las capacidades de computación que hemos experimentado en los últimos años y, en particular, gracias al *cloud computing* y las GPU, los investigadores han vuelto a retomar el modelo conexionista, ahora bajo el nombre de *deep learning*, que significa aprendizaje profundo. Este modelo está representado por una mastodónica red neuronal compuesta por millones de neuronas que es entrenada de forma supervisada para que aprenda ella misma a solucionar problemas concretos. En este caso, los investigadores ya se han resignado a no comprender cuáles son los procesos cognitivos de un sistema de *deep learning* cuando está en pleno funcionamiento.

El otro problema, el de la configuración, también queda rematado. En el momento en el que se construye una red neuronal de dimensiones astronómicas, la configuración que esta tenga es prácticamente irrelevante. El *deep learning*, también popularizado por la compañía Google, es hoy en día el motor de inteligencia artificial que existe dentro de gran parte de los sistemas de *big data*.

LA INTELIGENCIA ARTIFICIAL EN EL TRANSPORTE

En las últimas décadas se han producido pasos de gigante en lo referente a la inteligencia artificial aplicada al transporte, tanto en la conducción de coches autónomos como en la aviónica y la aeronáutica. Todos sabemos que, hoy en día, casi cualquier avión de transporte de pasajeros es capaz de aterrizar de forma automática, siempre que las instalaciones radioeléctricas del aeropuerto lo permitan. ¿Es esto una forma de inteligencia artifi-

cial? Como ya comentamos en la introducción del libro, es inevitable introducir un sesgo antropocéntrico en la definición de la inteligencia artificial. De hecho, el mismo test de Turing es eminentemente antropocéntrico: si no se consigue imitar el comportamiento de un humano, no se es inteligente. Pero imaginemos ahora un programa de ordenador que sea capaz de sincronizar y coordinar un complejo sistema de motor de vuelo, de tener un sistema de decisiones autónomo de rutas aéreas en función de las necesidades del momento y de presentar una fiabilidad del 100%. ¿Diríamos que es inteligente? Probablemente sí, dado que podría superar un test de Turing particular de pilotos de avión. En cambio, ¿diríamos lo mismo de una libélula o de un halcón peregrino? La primera realiza vuelos transoceánicos de miles de kilómetros y el segundo desciende en picado a casi 400 km/h. Ambos están dotados de un complejo sistema de motor de vuelo altamente fiable. Sin duda podrían superar un test de Turing y ser considerados inteligentes.

Hoy en día la inteligencia artificial ha permitido crear complejas aeronaves no tripuladas, los omnipresentes drones, muy útiles tanto para uso civil como militar, que no solo son capaces de maniobrar en entornos realmente adversos, sino también de desarrollar acciones igualmente complejas. Por ejemplo, detectar plagas en campos de cultivo, actuar como sistemas de vigilancia de fronteras o colaborar en las tareas de extinción de incendios forestales. Por supuesto, todo esto no sería posible si no fuera por el crecimiento exponencial de las capacidades de computación que permite integrar en estas aeronaves potentes entornos de cálculo.

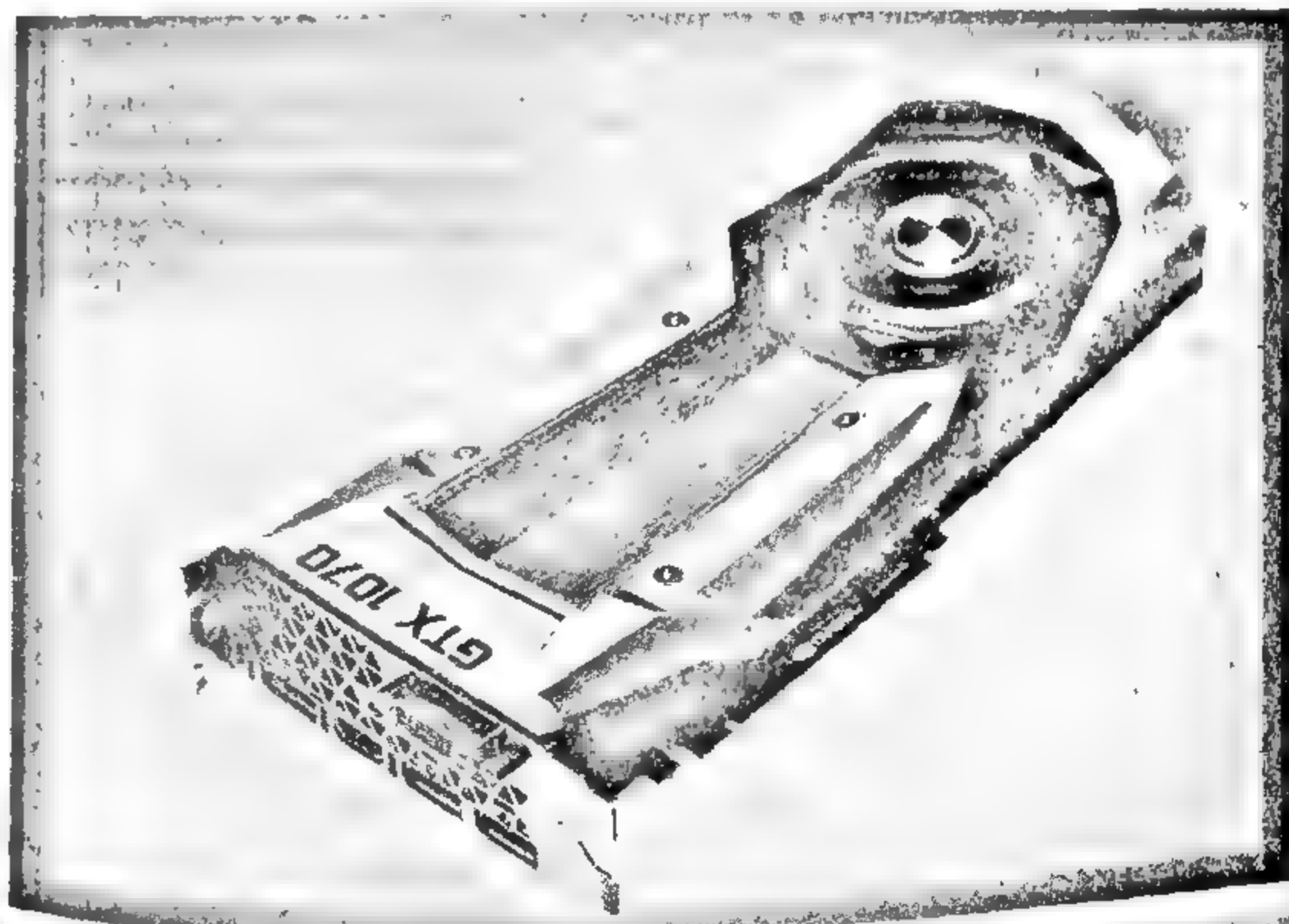
Una cosa similar sucede con los sistemas de conducción autónoma. La industria empezó ofreciendo al mercado «simpáticos» coches capaces de aparcar solos y ha continuado experimentando con vehículos que, de media, solo presentan un accidente por cada 100 millones de millas viajadas (o lo que es lo mismo, casi 161 millones de kilómetros), un hecho completamente inaudito en la conducción humana. Por tanto, y aunque resulte paradójico, estos sistemas no pasarían un test de Turing por ser demasiado infalibles.

LAS GPU: PROCESADORES DE GRÁFICOS

Las GPU o *Graphical Processing Units* son un tipo de procesadores muy especializados en el procesamiento de gráficos. Los que se usan en los videojuegos, por ejemplo, requieren un gran conjunto de cálculos matriciales para poder rotarlos, escalarlos o trasladarlos. Si una videoconsola tipo Sony PlayStation quiere tener unos gráficos realistas que gusten a los usuarios, necesita usar un hardware especial para poder realizar todas estas operaciones matemáticas en cuestión de fracciones de milisegundo. En caso contrario, los videojuegos no tendrían el realismo del que gozan hoy. Para conseguirlo, los ingenieros han desarrollado las GPU. Estos procesadores son realmente lentos en el cálculo de otro tipo de operaciones matemáticas, pero en lo que a operaciones matriciales se refiere, son imbatibles. Dada la gran evolución en los últimos años tanto de las videoconsolas como de los videojuegos, las GPU se han populanzado, abaratado y mejorado de tal manera que hoy en día disponen de miles y miles de nodos de cálculo en poco más de lo que ocupa la palma de la mano y por unos pocos centenares de euros.

Del ocio al cálculo complejo

Aunque las GPU fueron desarrolladas en el ámbito de la industria del videojuego para la aceleración de gráficos, pronto trascendieron a otras áreas de la informática. En la actualidad siempre que existe un proceso de cálculo costoso que contiene muchas operaciones matriciales, se intenta usar GPU de forma preferente. Sin duda el *deep learning* ha sido una de las áreas de la ingeniería que más se ha visto beneficiada por esta tecnología microelectrónica.



Pero no hace falta irnos a sistemas de conducción autónoma para poder hablar de inteligencia artificial en los coches. En la actualidad existe un gran número de modelos que incluso son capaces de detectar peatones mediante complejos sistemas de visión artificial y actuar en consecuencia, modificando la marcha si es preciso, dejando al margen las decisiones del conductor.

LA INTELIGENCIA ARTIFICIAL EN LA VIDA COTIDIANA

Es innegable que la inteligencia artificial nos rodea en todos los actos cotidianos que realizamos en nuestro día a día. Existen innumerables aplicaciones en este marco que no dejan de sorprendernos y que son usadas continuamente, desde sistemas inteligentes que escriben de manera automática textos en los periódicos —por ejemplo, noticias acerca de partidos de fútbol o sobre la evolución diaria de una cotización bursátil— hasta vehículos que circulan de forma autónoma, tal y como acabamos de ver. Lo más interesante de todo esto es que algunas de las grandes compañías que investigan intensivamente en este ámbito, como Google, Microsoft, IBM, etc., han hecho públicas algunas de sus tecnologías para que los ciudadanos de todo el mundo puedan experimentar con ellas. Como la que Google publicó gratuitamente en el año 2015, la tecnología TensorFlow, una «librería» de programación que permite aplicar *deep learning* a todo lo que queramos. En 2016, unos investigadores programaron TensorFlow para que aprendiera textos de Bob Dylan y poder así componer versos de forma automática con el mismo estilo que el premio Nobel de Literatura de 2016. Estos mismos ingenieros, en un derroche de imaginación, usaron la misma tecnología para que fuera posible expresarse como lo hacía Ramon Llull. No deja de ser irónico que un sabio que flirteaba con la idea de una inteligencia artificial prematura hace más de siete siglos, ahora sea imitado por una máquina inteligente.

Dejando al margen estas aplicaciones bizarras, la gran revolución de la inteligencia artificial tiene mucho que ver, como ya hemos dicho, con el enorme crecimiento de la capacidad de

cómputo, pero también con el avance de la sensorica. Ya hemos apuntado cómo los ingenieros han ido reproduciendo, uno a uno, los sentidos humanos de forma electrónica. De hecho, un teléfono móvil de gama media incorpora una quincena de sensores, más de los que tenemos los humanos, que son nueve. Con toda esta red sensorica y dejando de lado las definiciones más antropocéntricas de la inteligencia artificial, podemos afirmar que la inteligencia nos rodea 360° mediante una sarta de dispositivos inteligentes: teléfonos, televisores, electrodomésticos, las webs que visitamos cada día, e incluso coches «inteligentes». La combinación, por tanto, de los datos con el *cloud computing* y la sensorica ha permitido que la inteligencia artificial se haya introducido en todas nuestras vidas. Lo ha hecho de forma silenciosa, pero dando un gran valor añadido.

EL GO, LA PENÚLTIMA FRONTERA

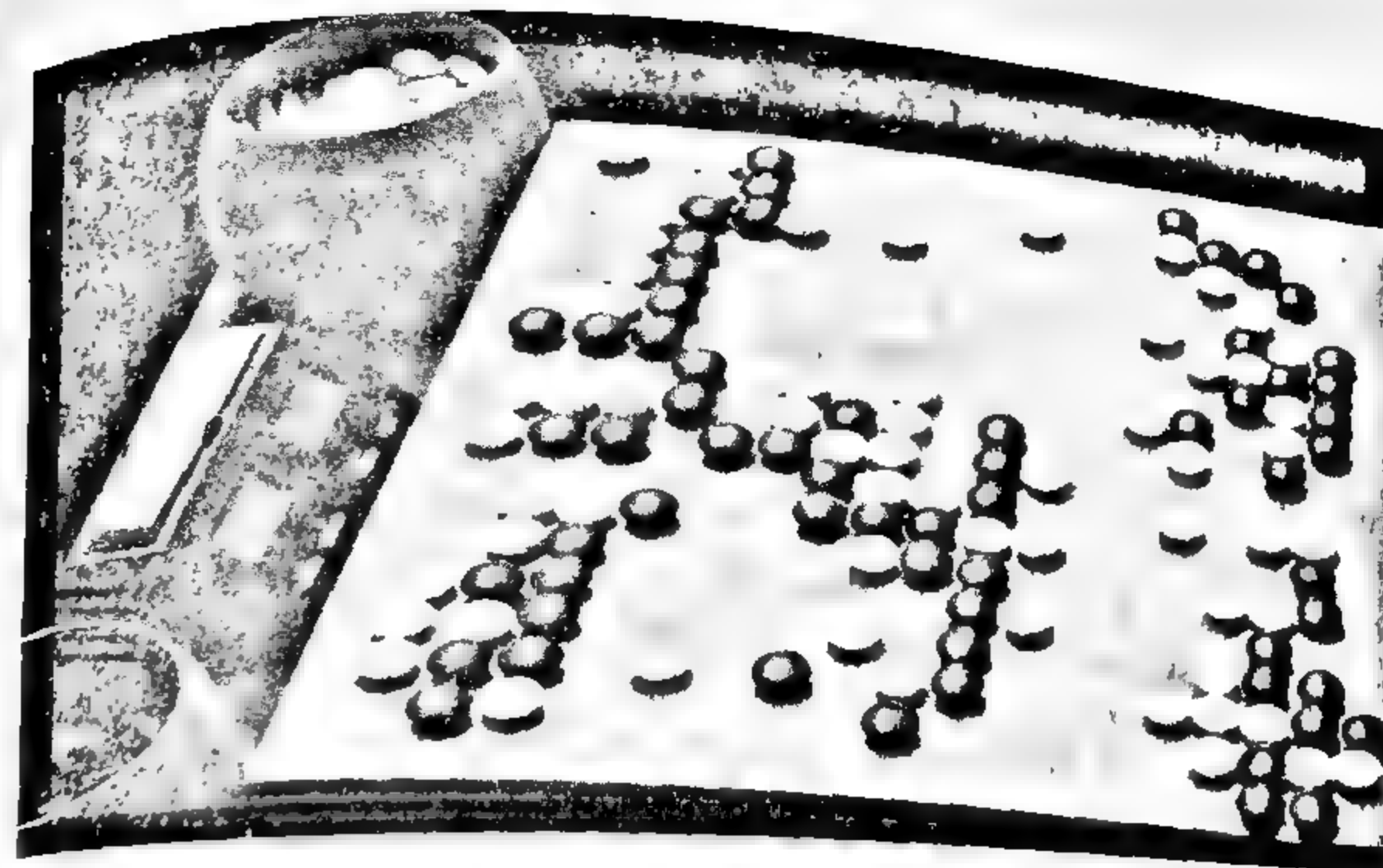
No podemos cerrar el presente capítulo sin hablar del go, ese juego de estrategia tradicional chino consistente en tratar de rodear al oponente mediante unas fichas blancas o negras. El go es un buen ejemplo de problema combinatorio frente al cual un humano medianamente entrenado puede detectar de un vistazo la estrategia más inteligente para cada escenario. Sin embargo, para un ordenador es extremadamente complejo. La importancia de este juego en la inteligencia artificial es crucial, ya que, hasta hace realmente poco, ningún programa de ordenador había sido capaz de vencer a ningún jugador de nivel profesional. Por tanto, crear sistemas de inteligencia artificial capaces de competir en go ha sido otro de los retos e hitos de la inteligencia artificial.

En este juego las reglas son extremadamente sencillas, pero dan lugar a escenarios de gran complejidad estratégica. Se trata de un tablero con una cuadrícula de 19×19 líneas donde dos jugadores van ubicando, en turnos alternativos, fichas blancas y negras en las intersecciones libres (véase la imagen superior de la pág. 123). Si una ficha o un grupo de ellas queda completamente rodeado por las del color contrario, el grupo queda capturado

y estas fichas se retiran del tablero. Se puede pasar turno si se cree conveniente, pero si los dos jugadores pasan consecutivamente, la partida se acaba y gana aquel jugador que en ese momento esté dominando una mayor porción del tablero de juego.

Matemáticamente, el go se clasifica como un juego de estrategia muy similar al ajedrez. Sin embargo, mientras que hay programas informáticos que son capaces de vencer de manera incontestable a los campeones mundiales de ajedrez —por ejemplo el programa Deep Fritz pudo vencer al campeón mundial Vladimir Kramnik por 6 a 0, en el año 2006, siendo ejecutado desde un ordenador portátil—, hasta hace muy poco ningún programa de go podía vencer a un jugador *amateur*. De hecho, el máximo hito al que se había llegado en este campo fue en el año 2008, cuando el programa Mogo pudo vencer a un jugador profesional. Sin embargo, este último tenía un hándicap de nueve fichas, es decir, Mogo pudo ubicar nueve fichas iniciales en el tablero en su primera tirada, y eso le proporcionó una enorme ventaja inicial.

Hay tres razones principales por lo que esto sucede. Una, que las dimensiones del tablero de go son más de cinco veces mayores que las del ajedrez, es decir 361 frente a 64, lo que implica un mayor número de jugadas para analizar. Dicho con otras palabras, en el ajedrez, el número de jugadas promedio ante cualquier posición de las piezas es de 37, mientras que en el go este número puede estar entre 150 y 250 por turno. El segundo motivo es que un movimiento de go puede tener influencia centenares de turnos después, siendo casi imposible para un ordenador hacer predicciones a tan largo plazo. Y, finalmente, en el ajedrez se puede evaluar con bastante precisión los beneficios que dará una jugada, ya que hay pocas piezas, se capturan de una en una y todas tienen un valor determinado. En el go, el beneficio que se puede obtener de una captura depende por completo del número de piezas cercadas, lo que depende de la situación de las piezas en aquel momento, siendo esta situación difícil de prever al depender del movimiento de piezas del contrario. Por tanto, para estimar el beneficio de una captura, se necesitan complejos cálculos estimatorios que, además, dan resultados relativamente poco fiables.



El go se juega en una cuadrícula de 19×19 líneas. Las fichas, llamadas «piedras», se colocan en sus intersecciones, como se observa en la fotografía superior. Abajo, partida de go jugada el 12 de marzo de 2016 en Seúl, Corea del Sur, por Lee Se-dol (derecha) contra Aja Huang, el programador de Google que maneja el programa AlphaGo.

Sin embargo, a lo largo de 2015 Google desarrolló el programa AlphaGo, basado en *deep learning*, con la intención de vencer a un jugador profesional. De esta manera, en marzo de 2016, AlphaGo se enfrentó al campeón Fan Hui en una serie de cinco partidas, que se decantaron a favor de la máquina por 3 a 2. Poco después, en el mismo mes, AlphaGo se enfrentó a otro campeón de aún más enjundia que el anterior, Lee Se-dol, quedando la disputa 4 a 1 a favor de AlphaGo (véase la fotografía inferior de la página anterior). El premio para los creadores de este programa fue de un millón de dólares que, como no podía ser de otra manera, Google donó a una institución benéfica, en este caso a Unicef. Por tanto, con AlphaGo y otros programas de go que han aparecido *a posteriori*, la que durante mucho tiempo fue una frontera de la inteligencia artificial también ha quedado superada.

CAPÍTULO 6

Algunas disquisiciones desde la ética y la filosofía

No es posible obviar los aspectos ligados a la filosofía y a la ética desde los cuales se abordan las distintas implicaciones de la inteligencia artificial. Sin duda existe un desencuentro conceptual recurrente entre cómo se diseñan esas máquinas y cómo las percibe la sociedad.

Ya iniciamos este libro presentando el debate existente entre la inteligencia artificial y el antropocentrismo. Para algunos, las aportaciones, la evolución y el estado actual de la inteligencia artificial es un proceso agri dulce, sobre todo porque las grandes expectativas creadas por todas aquellas películas y libros de ciencia ficción que, desde hace más de sesenta años, representan la inteligencia artificial como una tecnología que nos debe aportar robots humanoides o, en su defecto, máquinas como HAL 9000 —con toques tan humanos como la creatividad, la consciencia o los sentimientos— no se han cumplido.

Sin embargo, los ingenieros afirman que hoy la inteligencia artificial ya es omnipresente y que nos rodea en todas las acciones que realizamos a diario. Ciertamente es que no se trata de una inteligencia humanoide, pues precisamente viene a suplir las lagunas que el intelecto humano posee. Un ejemplo: los humanos no somos capaces de procesar mucha de la información que contienen los millones de pequeñas acciones y decisiones que cada uno de nosotros toma diariamente. En cambio, con todo este volumen de información y usando las técnicas de *big data*, la inteligencia artificial sí puede inferir reglas que escapan a la

capacidad de análisis de una mente humana. Por tanto, vistos estos dos puntos de vista, ¿dónde está el desencuentro entre los ingenieros y la sociedad? Sin duda alguna en la concepción de inteligencia artificial que tienen unos y otros. Mientras que los ingenieros la definen como una serie de tecnologías que prolongan la inteligencia natural hasta límites insospechados, la sociedad, en general, la imagina de tipo humanoide y, por tanto, antropocéntrica. Pero ¿qué es lo que define una inteligencia antropocéntrica? ¿Es la capacidad de expresar emociones y, por tanto, de sentirlas? ¿O, simplemente, hemos de circunscribir la inteligencia antropocéntrica a un mero juego de la imitación, tal y como lo definía Turing?

Como ya sabemos, Turing definió su famoso test como un método que permitía discernir si una máquina era inteligente o no en función de si un humano la podía confundir con otro humano en una conversación cualquiera. A pesar de las críticas tecnológicas y filosóficas del método, es el único test que, hasta hace muy poco, era universalmente aceptado para definir la inteligencia. Sin embargo, en los últimos años han aparecido otras alternativas, como es el caso del test del café, que define que un sistema inteligente es aquel capaz de preparar café en una casa familiar cualquiera. Es decir, un sistema capaz de encontrar el agua, el café, una taza, la cafetera, entender su mecanismo básico sin utilizar ningún manual técnico y, finalmente, preparar una buena taza de café.

Otro test de inteligencia es el del estudiante universitario. En él, un sistema se considerará inteligente si es capaz de matricularse en una universidad, tomar las clases que cualquier alumno debería atender y, por supuesto, licenciarse. O también podemos encontrar el test del empleo, es decir, un test que define un sistema como inteligente si es capaz de imitar el comportamiento de un humano en un trabajo de relevancia económica. En definitiva, todos ellos definen la inteligencia mediante pruebas físicas o intelectuales que millones de humanos realizamos cada día con toda naturalidad. No se ha definido formalmente otro tipo de test que determine la inteligencia a través de otros parámetros más internos, como podrían ser los sentimientos o la autoconsciencia.

De aquí la importancia trascendente de la película *Eva*, del año 2011, y múltiplemente galardonada. En dicha película surge una relación amorosa entre un humano y un robot humanoide. Lógicamente, el humano no conoce que se trata de una máquina hasta el último minuto de la película, en el que se desvela su identidad, con lo cual el robot no solo estaría pasando holgadamente un hipotético test de Turing, sino que, además, estaría consiguiendo que un humano lo amara. Por tanto ¿podríamos definir un «test de Eva», en el que la inteligencia se describe por la capacidad de generar lazos amorosos entre un humano y una máquina? ¿Por qué no?

Por otra parte, no hay duda de que ninguna tecnología es capaz, hoy en día, de crear consciencia, es decir, de desarrollar la facultad de decidir, hacerse sujeto y ser actor de sus actos y responsable de las subsiguientes consecuencias, según la percepción del bien y del mal. Difícilmente una máquina podría declarar, de forma espontánea, el «pienso, luego existo» de Descartes y, a partir de aquí, construir su propio sistema moral. Además, la forma interna de procesamiento de la información que usan muchos algoritmos considerados inteligentes podría encajar perfectamente con algunas definiciones filosóficas de la consciencia. Por ejemplo, la filosofía dialéctica define la consciencia como la reproducción por parte del individuo de la imagen ideal de su actividad tendente a una finalidad y de la representación ideal en ella de las posiciones de las otras personas. Siguiendo esta definición, podríamos afirmar que tecnologías como Deep Blue están dotadas de consciencia.

LA FILOSOFÍA Y LA INTELIGENCIA ARTIFICIAL

La inteligencia artificial y la filosofía son dos disciplinas que siempre han ido de la mano. En este libro hemos ido introduciendo el pensamiento de filósofos de la talla de Aristóteles y Ramon Llull, en lo que hemos llamado la «prehistoria de la inteligencia artificial», y más adelante, el de otros personajes relevantes, como Turing —que también es definido como un filósofo— o

La amenaza de la humanidad
no son las máquinas, son los
hombres.

RAÚL ROJAS

Wittgenstein. La misión de la filosofía es responder a preguntas trascendentales que los humanos nos solemos hacer y que incluyen cuestiones como ¿puede una máquina actuar inteligentemente?, ¿son la inteligencia natural y la artificial una misma cosa?, ¿puede una máquina tener consciencia? A todas estas preguntas los filósofos han ido dando respuesta. Según Turing, si una máquina puede comportarse tan inteligentemente como un humano, entonces es que es tan inteligente como él. Y, según Searle, un computador programado apropiadamente, con las adecuadas entradas y salidas, tendría una mente en el mismo sentido que la tiene un ser humano. Todas estas respuestas, desde luego, tienen un gran trasfondo filosófico.

Sin duda alguna, la mayor parte de los esfuerzos filosóficos se centran en la pregunta de si una máquina puede tener una inteligencia equivalente a la humana. Como hemos ido viendo desde Turing, las reflexiones y posturas ante esta cuestión han sido muy variadas. Probablemente la respuesta más protecnológica es la que se dio como resumen de la conferencia de Dartmouth: cada uno de los aspectos del aprendizaje u otras de las características de la inteligencia puede ser definido de una forma tan precisa que siempre se podrá construir una máquina para simularlo. Lógicamente, esta premisa tiene sus defensores y sus detractores. Veamos algunas de las posturas más destacables.

El filósofo contemporáneo estadounidense Hubert Dreyfus (n. 1929) opina que, si el sistema nervioso obedece a las leyes de la física y la química, entonces debería poderse imitar con un dispositivo físico diseñado para tal propósito. La principal crítica a este planteamiento es que las neuronas que componen un sistema nervioso, que suman 10^{11} , son tantas que es extremadamente difícil simularlas de forma artificial. De hecho, en el año 2005 se consiguió simular un segundo de pensamiento de un sistema neuronal de 10^{11} . Esta simulación tardó en realizarse 50 días en un supercomputador de 27 procesadores. En otras palabras, podríamos decir que se tardaron 50 días en simular el

pensamiento de lo equivalente a un segundo de un bebé recién nacido. Lógicamente, si tuviéramos más potencia de cálculo a nuestra disposición, este tipo de simulaciones serían más factibles. Quizá, cuando tengamos acceso a computadores cuánticos de propósito general, esto ya será posible. ¿Quién sabe?

Otro argumento es el que expresaron los científicos Newell y Simon, los cuales afirmaban que un sistema simbólico tiene los medios necesarios y suficientes para construir una acción de inteligencia general. Es decir, que la inteligencia humana no es más que un complejo sistema de manipulación de símbolos que nos permite pensar, comunicarnos, tener creatividad, consciencia, etc. Por tanto, un computador que implemente un sistema simbólico es inteligente por sí mismo. Este argumento, como era de esperar, despierta muchas reacciones en contra. Entre ellas las de los que afirman que un sistema simbólico es necesario pero no suficiente para crear inteligencia, ya que esa facultad es mucho más que una mera manipulación ordenada de símbolos.

Un tercer razonamiento es la tesis antimecanicista de Gödel. El filósofo y matemático austriaco-estadounidense Kurt Gödel (1906-1978) desarrolló en 1931, con solo veinticinco años de edad, sus dos célebres teoremas de incompletitud, los cuales demuestran que, bajo ciertas condiciones, ningún sistema de lógica formal puede ser consistente y completo a la vez. Según los expertos, estos teoremas constituyen uno de los grandes avances de la lógica matemática, y su repercusión en las ciencias de la computación ha sido trascendental. Ambos teoremas de la incompletitud están relacionados con la existencia de proposiciones indecidibles en ciertas teorías aritméticas. El primero nos dice que, bajo ciertas condiciones, ningún sistema matemático formal capaz de describir los números naturales y la aritmética con suficiente expresividad es a la vez consistente y completo. Es decir, si los axiomas de dicha teoría no se contradicen entre sí, entonces existen enunciados que no pueden probarse ni refutarse a partir de ellos. El segundo es un caso particular del primero y demuestra que una de las sentencias indecidibles de dicho sistema matemático formal es aquella que «afirma» la

consistencia del mismo. O lo que es lo mismo: si el sistema de axiomas en cuestión es consistente, no es posible demostrarlo mediante dichos axiomas. Eso viene a decir que siempre habrá axiomas que no puedan ser formalmente demostrados, aunque la mente humana los dé por válidos, lo que conlleva que el poder de la mente humana no puede ser reducido a un mecanismo. En otras palabras: el razonamiento humano no es computable. A pesar de esto, existen numerosos científicos, como Roger Penrose, que afirman y razonan de qué modo este escollo se superará mediante la computación cuántica.

Otro argumento de Dreyfus, en este caso en contra de la posibilidad de crear una inteligencia artificial, es el que defiende que la inteligencia humana depende primariamente de instintos inconscientes más que de la manipulación simbólica. Como estos instintos inconscientes nunca podrán ser capturados por un sistema de reglas formal, la inteligencia artificial no es posible. La gran crítica a este argumento viene de la mano de los conexionistas, que afirman que las aproximaciones *bottom-up* sí que pueden proporcionar algunos de estos instintos inconscientes.

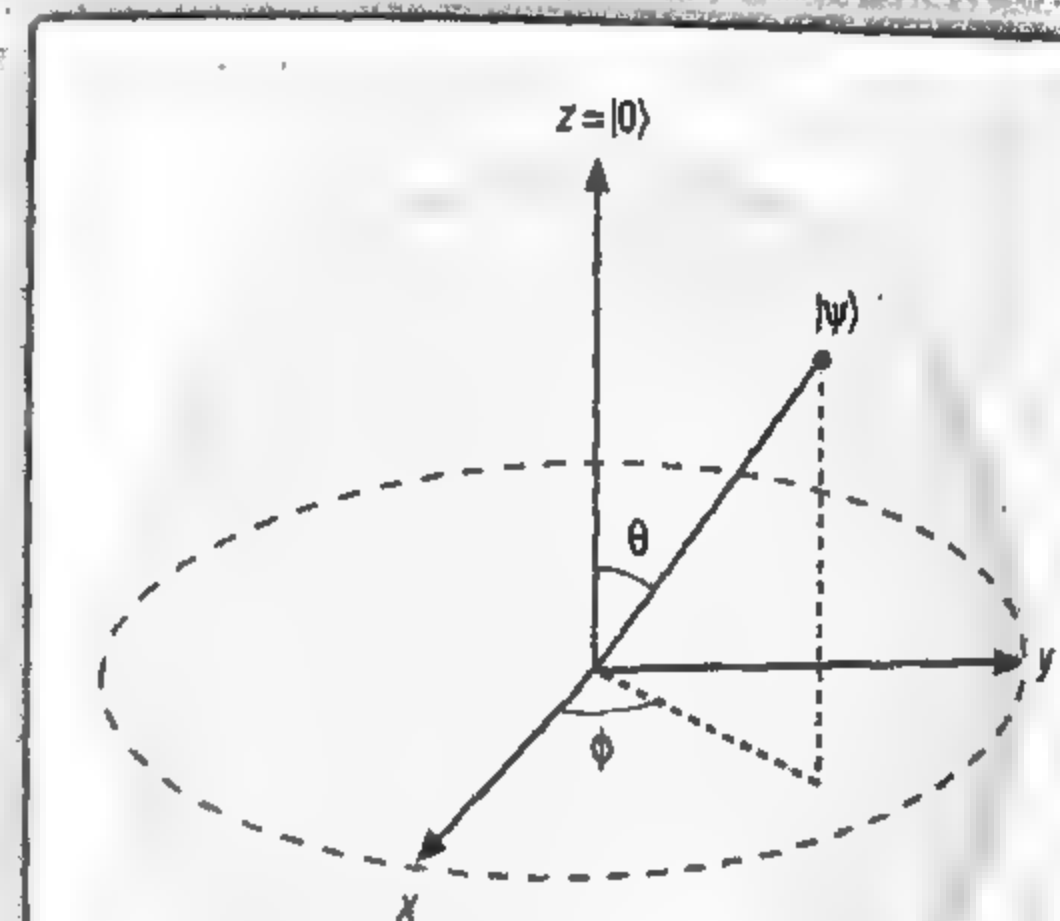
Turing proclamó que existen diversas afirmaciones que nunca podrían ser ciertas. Estas serían de la forma «una máquina nunca hará X», donde X puede ser: ser amable, bonito, amigable, tener iniciativa, tener sentido del humor, mentir, cometer errores, enamorarse, disfrutar de las fresas con nata, seducir a alguien, aprender de la experiencia, usar las palabras correctamente, ser el sujeto de su propio pensamiento, tener tanta diversidad de comportamientos como un humano o hacer algo realmente novedoso. A lo largo de este libro hemos visto cómo la inteligencia artificial ha avanzado lo suficiente como para que muchas de estas afirmaciones se hayan cumplido, y estamos en camino de lograr afirmar algunas otras. Por tanto, podemos decir que Turing no fue capaz de prever el desarrollo exponencial que tendría la tecnología en el futuro y, en particular, la inteligencia artificial. Así que, para no volver a caer en el mismo error, ahora los filósofos, científicos e ingenieros se deben formular preguntas del tipo: ¿podrán alguna vez las máquinas tener tal o cual cualidad humana? ¿Podrán sentir emociones, tener autoconsciencia, ser

LA COMPUTACIÓN CUÁNTICA

Los grandes avances cualitativos y cuantitativos en la inteligencia artificial siempre han venido precedidos de un salto tecnológico en las capacidades de computación. Sin duda, el próximo salto de este tipo es el que originará la computación cuántica, un nuevo paradigma de cálculo que deja de basarse en los bits para centrarse en los qubits. El qubit, término atribuido al físico teórico estadounidense Benjamin Schumacher, es la unidad mínima que nos ofrece un sistema cuántico y, por tanto, constitutiva de la teoría de la información cuántica.

Qubits: capacidad incalculable

La gran diferencia entre un qubit y un bit es la cantidad de información que pueden contener y la forma en que esta puede ser manipulada. Codificar la información mediante qubits ha permitido desarrollar nuevas puertas lógicas, es decir, operaciones lógicas implementadas electrónicamente, que no pueden funcionar en un sistema binario, de verdadero o falso, o de unos y ceros, como lo son los bits. Los dos estados básicos de un qubit son $|0\rangle$ y $|1\rangle$, que corresponden al 0 y al 1 del bit clásico. Pero, además, el qubit puede encontrarse en un estado de superposición cuántica, una combinación de esos dos estados. En esto es significativamente distinto al estado de un bit clásico, que puede tomar solamente los valores 0 o 1. Un qubit puede representarse mediante la denominada esfera de Bloch (véase la figura), que permite asociar cada punto sobre su superficie con cada estado posible del qubit. Se estima que los computadores cuánticos que en estos momentos se están desarrollando aportarán tal potencia de cálculo que muchos de los grandes problemas que hasta ahora se consideraban incalculables pasarán a ser resueltos. Algo que, desde luego, beneficiará, y mucho, al campo de la inteligencia artificial.



La esfera de Bloch es una representación de un qubit que, en mecánica cuántica, representa geométricamente el espacio de estados puros de un sistema cuántico de dos niveles.

originales o creativas, mostrar hostilidad o benevolencia e, incluso, tener alma? Veamos las respuestas más extendidas a algunas de estas preguntas.

SENSIBILIDADES Y APTITUDES MECANIZADAS

¿Pueden tener las máquinas emociones? En primer lugar cabría preguntarse cuál es la definición de emoción, dado que la respuesta puede ser diversa, en función de cómo entendamos qué es una emoción. En segundo lugar hay que plantearse si realmente podemos decir que un sistema que imita tener emociones genuinas, hasta el punto de engañar a cualquiera, realmente las tiene o son simples simulaciones. Una definición que se suele usar al respecto es que una emoción es una acción que influye en un patrón de comportamiento. Por ejemplo, el miedo es una emoción que influye en un desarrollo urgente de las subsiguientes acciones, y la empatía es una emoción que optimiza una experiencia de interacción con un humano. Incluso las emociones como el amor pueden ser modeladas por los ingenieros, siendo un patrón de comportamiento que conlleva a quien lo siente a complacer al otro de una forma completamente altruista, dado que de esta forma se obtiene un refuerzo positivo.

Ascendamos ahora al siguiente nivel, el de la autoconsciencia. ¿Puede ser una máquina autoconsciente? Por el momento, como ya hemos dicho, hoy en día ninguna tecnología ha demostrado tener consciencia. Pero la respuesta podría ser diferente, de nuevo, según cómo definamos el concepto de la autoconsciencia. Si consiste en pensar sobre sí mismo, una máquina, sea inteligente o no, puede ser autoconsciente. Por ejemplo, cualquier coche tiene indicadores internos sobre el funcionamiento del motor, niveles de aceite, etc.; si alguno de estos indicadores se dispara, por motivos de seguridad el coche puede llegar a pararse. Por tanto ¿podemos decir que este coche es autoconsciente?

Pero ¿y qué pasa con la creatividad u originalidad? Turing define la originalidad de una máquina como la capacidad de sorprendernos. Sin duda alguna, este concepto hace muchas décadas



Existen algoritmos, como el desarrollado en la Universidad de Tübinga, que «repintan» una fotografía imitando el estilo de cualquier pintor. Las imágenes muestran una escena de la ciudad de Tübinga reproducida emulando a Van Gogh. Compárese con *La noche estrellada* (abajo a la izquierda), una de las obras más conocidas del artista neerlandés.

que está superado. Un ejemplo: la computación evolutiva, una de las técnicas de la inteligencia artificial, es capaz de diseñar nuevas estructuras de ingeniería civil que sorprenden hasta a los más expertos, y también de componer originales piezas musicales. En la introducción del libro hablamos de un sistema capaz de imitar el estilo de Van Gogh de forma muy creativa y también el de otros grandes pintores (véase la imagen de la página anterior). Ahora podemos decir que este procedimiento, publicado por la Universidad de Tubinga, está basado en el *deep learning*.

Emociones, autoconsciencia, creatividad... Podríamos seguir elaborando una larga lista de propiedades eminentemente humanas hasta aburrirnos. La respuesta, por lo general, siempre va a ser que una máquina puede adquirir tal o cual propiedad, siempre que esta sea definida de una forma muy descriptiva, racional y conductual.

Por tanto, la última de las preguntas sobre las que cabe reflexionar es ¿podrán tener algún día las máquinas alma? Sin duda, a lo largo de los siglos, este ha sido el caballo de batalla para definir al ser humano. ¿Es el alma lo que distingue a las personas de los demás animales? En caso afirmativo, ¿podremos algún día crear máquinas con alma? Turing definió la objeción teológica como la opinión que proclama que pensar es una función del alma inmortal. En este contexto, Turing contraargumentó:

Al tratar de construir tales máquinas, no deberíamos estar irreverentemente usurpando su poder de crear almas, como tampoco lo es en la procreación de los niños; más bien, en cualquier caso, somos instrumentos de Su voluntad que proporcionan mansiones para las almas que Él crea.

YO, ROBOT ÉTICO Y MORAL

En su relato *Círculo vicioso*, de 1941, el escritor de ciencia ficción Isaac Asimov introdujo por primera vez tres leyes de la robótica, que más adelante se extendieron a cuatro (la cero es la que se introdujo *a posteriori*); dichas leyes son:

— Un robot no herirá a un ser humano o, por inacción, permitirá que un ser humano sea dañado.

— Un robot debe obedecer las órdenes dadas por un ser humano excepto cuando estas entren en conflicto con la primera ley.

— Un robot debe proteger su propia existencia, siempre y cuando dicha protección no entre en conflicto con las leyes primera y segunda.

— Un robot no dañará a la humanidad o, por inacción, permitirá que esta sea dañada.

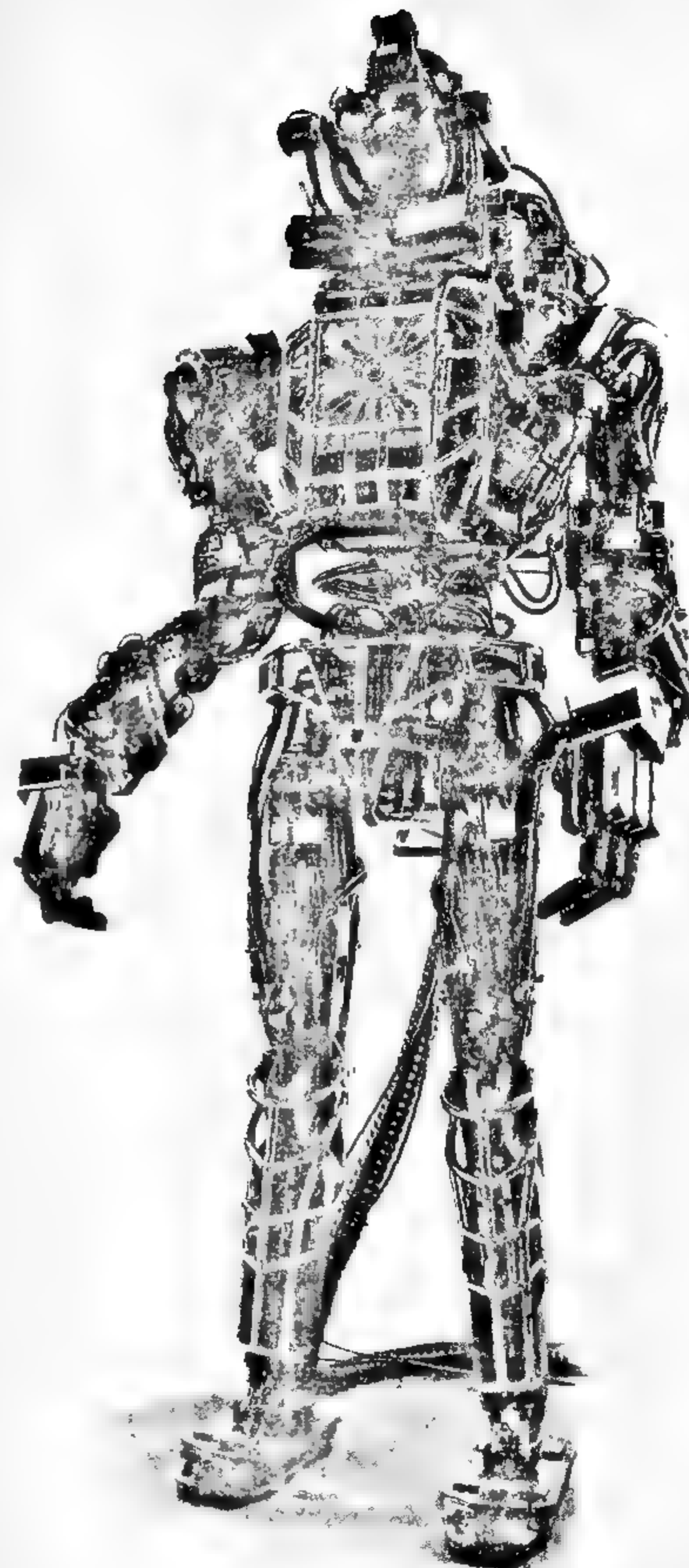
Si en la década de 1940 estas leyes eran pura ciencia ficción, hoy en día están más vigentes que nunca, sobre todo como una guía básica para el desarrollo ético y moral de la inteligencia artificial. De hecho, en los últimos años se han sucedido diversos experimentos con entes inteligentes y se ha observado que, en determinadas circunstancias, dichos entes han aprendido a hacer trampa para maximizar su retorno. Por ejemplo, en la École Polytechnique Fédérale de Lausana, en Suiza, se programaron diversos robots para que cooperaran en la tarea de encontrar recursos energéticos y tratar de evitar recursos «peligrosos». El conjunto de robots empezó a cooperar correctamente, pero con el tiempo esta cooperación degeneró en la popularización de diversas trampas entre ellos para maximizar el retorno egoísta. Otro ejemplo es el programa inteligente diseñado para aprender a jugar al Tetris, el cual desarrolló de forma autónoma una función que era pulsar el botón de pausa de forma indefinida cuando estaba a punto de perder. De esta forma no perdía nunca y maximizaba su retorno. Obviamente, estamos ante ejemplos relativamente inofensivos, pero sirven de advertencia para que los ingenieros lo tengan en cuenta y evitar así males mayores en otro tipo de inteligencia artificial, como por ejemplo agentes inteligentes con funciones militares, de cuidado de personas o a cargo de tareas críticas. No en vano, Google ha desarrollado lo

que ellos llaman el *kill switch*, es decir, un botón que desactiva a un agente inteligente en caso que su comportamiento no esté siendo el correcto.

Por todos estos motivos, existen diversos autores y pensadores que proclaman los graves peligros a los que se expone la humanidad si se continúan desarrollando máquinas inteligentes (en la página contigua, un modelo de la empresa de Google Boston Dynamics), desde la extinción de nuestra especie hasta los estragos militares que puede causar una super inteligencia artificial en manos equivocadas. En este sentido, solo hace falta citar las palabras del célebre físico Stephen Hawking, una de las voces más enérgicas contra este tipo de tecnología: si bien el desarrollo de la inteligencia artificial podría ser el evento más grande en la historia humana, por desgracia, también podría ser el último.

Todo este debate ético sobre la inteligencia artificial está hoy a la orden del día con motivo del advenimiento del coche de conducción autónoma. El dilema moral recurrente en esta tecnología es el siguiente: imaginemos que nuestro coche autónomo circula por una carretera por donde han cruzado imprudentemente, sin mirar, cinco personas. Si el coche continúa, esas personas van a ser atropelladas con consecuencias fatales. Sin embargo, si el coche gira en el último momento, atropellará a otra persona, pero solo una, que está cruzando por otra calle donde el semáforo de peatones está en verde. ¿Qué tiene que hacer el coche? Visto desde el punto de vista más frío y racional, la respuesta es obvia: la decisión que tiene que tomarse es la que menos daño haga, es decir, atropellar a una sola persona en vez de a cinco. Pero el dilema moral, desde luego, está servido, ya que esta persona que va a ser atropellada ha cruzado cuando debía, contrariamente a los otros cinco que se han saltado su semáforo. ¿Es ético atropellar a una persona que ha actuado correctamente para salvar a cinco que no lo han hecho?

Ahora imaginemos esta misma situación, pero donde la única forma de salvar a las cinco personas que cruzan imprudentemente es dando un brusco giro que lance al propio coche y a su



Este robot humanoide llamado Atlas fue presentado en 2013 por la empresa Boston Dynamics en el concurso de robótica organizado por la Agencia estadounidense de Proyectos de Investigación Avanzados de Defensa, DARPA. Con 1,8 m de altura, puede operar en terrenos accidentados, escalar y transportar objetos y su cabeza está equipada con sensores, cámaras y un localizador láser.

ocupante, una sola persona, por un precipicio. De esta forma, se sacrifica una vida, la del ocupante del coche, en pro de otras cinco vidas. ¿Alguien se subiría en un coche capaz de tomar estas decisiones? No parece probable...

Una inteligencia artificial aún más inteligente es una de las amenazas más graves del siglo XXI.

YUVAL NOAH HARARI

Al menos, exigen algunos pensadores, deberíamos forzar a que los ingenieros implementen alguna suerte de caja negra en las máquinas inteligentes, de tal forma que cuando suceda algún evento fatal podamos entender el porqué de esas decisiones tomadas en tiempo real. En otras palabras, la exigencia moral de implementar cajas negras en los agentes inteligentes permitirá saber si una máquina ha «enloquecido» o, si por el contrario, ha tomado la mejor decisión racional que, con la información de la que disponía en ese momento, se podía tomar.

¿QUE VIENE EL FUTURO!

Ante todos estos desafíos, la administración Obama decidió encargar, en octubre de 2016, un informe sobre la inteligencia artificial y sus retos a un grupo de expertos en la materia. Este informe se ha hecho público y está a disposición de cualquier ciudadano para su consulta, lo cual es de agradecer.

En primer lugar, el informe responde a la pregunta de qué es la inteligencia artificial. Los expertos dan cuatro respuestas válidas, síntoma de que no existe una única respuesta consensuada. Estas fueron:

1. Sistemas que piensan como humanos, como, por ejemplo, arquitecturas cognitivas o redes neuronales.
2. Sistemas que actúan como humanos y que, por tanto, superan el test de Turing mediante el procesamiento del lenguaje natural, sistemas de representación del conocimiento y razonamiento y aprendizaje artificial.

3. Sistemas que piensan racionalmente, como sistemas de optimización o inferencia.

4. Sistemas que actúan racionalmente, como agentes inteligentes o robots que implementan sistemas perceptivos, de planificación, de razonamiento, de aprendizaje, de comunicación y de toma de decisiones.

En el mismo informe se reportaba una encuesta a múltiples investigadores expertos en inteligencia artificial. Las dos conclusiones más significativas son que el 80% de los encuestados opina que la inteligencia artificial del nivel de un humano llegará tarde o temprano, y el 50% de estos considera que existe un 50% de probabilidades de que esto suceda alrededor del año 2040.

Por otra parte, estos mismos expertos afirman que los retos más importantes que se deben abordar en los próximos años son:

- Conseguir que las máquinas puedan realizar tareas más amplias y menos estructuradas.
- Unificar los diferentes «estilos» de inteligencia artificial, es decir, las históricas aproximaciones *bottom-up* y *top-down*, que todavía hoy en día evolucionan por caminos separados.
- Solucionar problemas técnicos específicos como, por ejemplo, la transferencia del aprendizaje.

Y yendo más al detalle, se han citado otros elementos más concretos que también deben ser resueltos como:

- Evitar efectos negativos: Recordemos lo que pasaba con los robots suizos o el sistema que aprendió a pausar el Tetris para no perder. Este tipo de comportamientos se pueden extrapolar a casi cualquier entorno; por ejemplo, si programamos una máquina de limpiar suelos para que su recompensa se mida por la cantidad de suciedad limpiada,

esta puede aprender a ir tirando jarrones u otros elementos decorativos para limpiar más.

- Evitar el *hackeo* de recompensas: continuando con el ejemplo del robot limpiador, podemos programar un desincentivo si hace «travesuras» como la de tirar un jarrón para simular que ha limpiado mucha cantidad de suciedad. Sin embargo, si tiene un buen sistema de aprendizaje, el robot podría aprender a desactivar su sistema de visión artificial para no detectar que ha hecho alguna «travesura». Ya se conoce el dicho: ojos que no ven, corazón que no siente.
- Supervisión escalable: los robots pueden encontrarse a veces ante complejos dilemas para completar sus tareas. Ante ellos, deben encontrar mecanismos rápidos, sencillos y escalables para actuar. Sin embargo, si el sistema de recompensa favorece demasiado la simplicidad en pro de la eficiencia, podemos volver a encontrarnos con decisiones que no nos gustan. Por ejemplo, ante un teléfono móvil tirado por el suelo, un robot de limpieza puede decidir tirarlo a la basura favoreciendo la simplicidad, sin antes consultar a un humano si debe o no desecharlo.
- Exploración segura: lógicamente, los robots y los agentes inteligentes en general tienen que saber explorar el medio. Sin embargo, esta exploración tiene que tener unos límites de seguridad que eviten males mayores. Por ejemplo, un robot de limpieza no debería verter agua ni otros productos líquidos sobre zonas que tienen tensión eléctrica.
- Robustez y cambio distribucional: los robots que son entrenados en un determinado ambiente tienen que tener la suficiente inteligencia para poder extrapolar su experiencia adquirida y trasladarla a entornos diferentes. Desde luego, si un robot de limpieza es entrenado en un entorno fabril, las precauciones que se deben tomar y que el robot habrá aprendido son muy diferentes de las precauciones

que deberán adoptarse en un hogar. Por ello, la inteligencia artificial tiene que proporcionar los mecanismos correspondientes para que esta traslación pueda ser realizada de forma correcta.

En definitiva, para aproximar todos estos retos, el Gobierno de Estados Unidos invirtió 1 100 millones de dólares en el desarrollo de la inteligencia artificial para uso civil —no se ha informado de la inversión para usos militares—. Con estos niveles de inversión, en máximos históricos, quién sabe si todos los retos citados no serán solucionados mucho antes del año 2040.

Y, más allá del momento en el que las máquinas lleguen al nivel de inteligencia de un humano, otro punto clave que se tiene que analizar es el momento conocido como *singularity* o singularidad tecnológica, que hace referencia al hipotético momento en que las máquinas no solo superen en inteligencia a los humanos, sino que, además, sean capaces de autorreplicarse y mejorarse en el proceso. ¿Verá nuestra generación ese momento?

LECTURAS RECOMENDADAS

- CASTI, J.L., *El quinteto de Cambridge*, Madrid, Taurus, 1998.
- COPELAND, J., *Inteligencia artificial*, Madrid, Alianza, 1996.
- FAVERÓN, O., *La inteligencia artificial de Google*, 2016.
- GARDNER, H., *La nueva ciencia de la mente. Historia de la revolución cognitiva*, Barcelona, Ediciones Paidós, 2004.
- HOFSTADTER, D.R., *Gödel, Escher, Bach. Un eterno y grácil bucle*, Barcelona, Tusquets, 2015.
- LAHOZ-BELTRA, R., *Turing. Del primer ordenador a la inteligencia artificial*, Madrid, Nivola, 2009.
- MCCORDUCK, P., *Máquinas que piensan: una incursión personal en la historia y las perspectivas de la inteligencia artificial*, Madrid, Tecnos, 1991.
- PALMA, J.T. Y MARÍN, R., *Inteligencia artificial. Técnicas, métodos y aplicaciones*, Madrid, McGraw-Hill/Interamericana de España, 2008.
- PENROSE, R., *La nueva mente del emperador*, Barcelona, DeBolsillo, 2015.
- SEARLE, J.R., *Mentes, cerebros y ciencia*, Madrid, Cátedra, 2001.
- SOMERS, H. Y HUTCHINS, W.J., *Introducción a la traducción automática*, Madrid, Visor Libros, 1995.

VIVANCOS, D., *Big Data. Hacia la inteligencia artificial*, Madrid, The Valley, 2016.

ÍNDICE

aeronáutica 117
ajedrez 32, 44, 62, 72, 73, 88, 90-93, 122
AlphaGo 123, 124
Amazon 32, 99, 100, 114
Apple 48, 85, 87
Aristóteles 15-17, 24, 25, 129
Ars Magna 22, 23
aviónica 117

Babbage, Charles 28-33, 37
big data 18, 96, 103, 105-107, 110-113, 116, 117, 127
Boole, George 22, 24, 28

cloud computing 97, 100, 112, 117, 121
coche autónomo 138
conexionista, modelo 46, 59, 63, 68, 75, 88, 111, 116, 117, 132
conferencia de Dartmouth 57, 58, 63, 130

Ctesibio de Alejandría 16, 17, 19

De Morgan, Augustus 24, 26-28
Deep Blue 44, 88, 90-93, 129
deep learning 111, 116, 117, 119, 120, 124, 136
dron 118

Eliza 44, 46

go 121-124
Google 100, 106-108, 110, 111, 113, 117, 120, 123, 124, 138
GPU 117, 119
grid computing 98, 99
Guerra Fría 55, 59, 70, 81

IBM 44, 85, 87, 90-92, 94, 96, 97, 120

Jeopardy! 44, 91

- Kaspárov, Garri 44, 88, 90-93
- Lisp Machine 85-87
- lógica 15-17, 23, 24, 26-28, 50, 59, 63, 68, 73, 79, 112, 131
- Lovelace, Ada 30, 43
- map-reduce, técnica del 111, 112
- máquina
- analítica 28-30
 - de soporte vectorial 78, 116, 117
 - de Turing 37, 45
- Massachusetts Institute of Technology (MIT) 57, 88, 89
- McCulloch, Warren 47, 63
- Minsky, Marvin 57, 66, 68, 71, 73, 78
- Musa Al-Juarismi 18, 20, 22
- Newell, Allen 26, 57-59, 61, 62, 66, 131
- no-separabilidad lineal 68, 70, 75, 78
- nube, computación en la 97-101
- Pitts, Walter 47, 63
- Ramon Llull 22-24, 120, 129
- redes neuronales 46, 47, 57, 64, 65, 74, 75, 77, 78, 80, 116, 140
- Rosenblatt, Frank 57, 59, 63, 65
- san Isidoro de Sevilla 18, 21
- Searle, John 42, 130
- sensorica 121
- Simon, Herbert 26, 57-59, 61, 62, 71, 131
- singularity (singularidad tecnológica) 143
- sistemas
- expertos 43, 74, 85, 86, 88
 - simbólicos 58-60, 62, 63, 70, 88
- sobreentrenamiento 76
- test de Turing 33, 38, 40-42, 44, 88, 90, 91, 118, 129, 140
- Torres Quevedo, Leonardo 30, 31, 33, 37, 90
- Turco, El 31, 32, 90
- Turing, Alan 11, 13, 30, 33, 37, 38, 40-42, 46-49, 52, 53, 57, 59, 91, 128-130, 132, 134, 136
- Watson 44, 91, 94-97
- Wittgenstein, Ludwig 47, 49, 50, 52, 53, 130